T ARA

UNIVERSITAT POLITÈCNICA DE CATALUNYA

Dep. d'Enginyeria de Sistemes, Automàtica i Informàtica Industrial

TESI DOCTORAL

APORTACIÓ ALS MÈTODES DE SEGUIMENT TRIDIMENSIONAL D'OBJECTES, D'ALTA VELOCITAT D'OPERACIÓ, MITJANÇANT L'ESTEREOVISIÓ.

Joan Aranda López

Director: Josep Amat i Girbau

Barcelona, setembre de 1997

UPC
UNIVERSITAT POLITÈCNICA DE CATALUNYA

Departament d'Enginyeria de Sistemes, Automàtica i Informàtica Industrial

Edifici U
c/. Pau Gargallo, 5
08028 Barcelona
Tel. (+34 3) 401 69 74
Fax (+34 3) 401 70 45

Tesi doctoral: **Aportació als mètodes de seguiment tridimensional d'objectes, d'alta velocitat d'operació, mitjançant l'estereovisió**

Doctorand: **Joan Aranda López**

Director: **Josep Amat Girbau**

Lloc de lectura: **Dept. d'Enginyeria de Sistemes, Automàtica i Informàtica Industrial**
Edifici U, Campus Sud, Barcelona

Data de Lectura: 16 d'octubre de 1997

Aquesta tesi ha estat enregistrada
amb el núm. 188

Als meus pares,
Juan i Justa

AGRAÏMENTS

Al Dr. Josep Amat, per haver acceptat la responsabilitat (amb la feina extra que ha comportat a la seva ja carregada agenda) de dirigir aquesta tesi.

A la resta de companys del departament d'ESAII, als que encara hi són i als que han passat per aquí, pels seus consells relatius a la tesi i el seu suport professional i també emotiu.

A aquells alumnes-col·laboradors que sense més interès que el d'aprendre m'han ajudat a implementar el sistema presentat. Especialment vull agrair l'ajut de José Juan Calvo i més recentment d'Helena Garrido, que gràcies a la seva il·lusió, esforç i dedicació han fet possible la implementació existent.

A la Dra. Carina Gibert, per deixar-se enredar en la revisió del capítol dedicat a l'anàlisi d'errors. Els seus comentaris i suggeriments han enriquit notablement el formalisme matemàtic de la tesi.

Finalment a la Rosa, per molt més que les hores dedicades a l'edició de figures per a la memòria.

A tots ells, només espero poder correspondre'ls.

ÍNDEX

SUMARI

Sumari

El sistema de seguiment tridimensional d'objectes que es presenta, pretén ser una aportació més a la recerca bàsica en Visió per Computador. No obstant, s'ha fet un gran esforç en la recerca d'una metodología que permeti la seva aplicabilitat en sistemes automatitzats que necessitin d'una realimentació en *temps real*. Amb aquest objectiu, s'ha implementat una tarja processadora d'imatge de baix cost (*'low cost'*) que incorporada a un PC estàndard, proporciona al sistema una elevada relació velocitat-cost, la qual cosa el pot fer molt atractiu a curt termini en *aplicacions industrials*.

La *transformació polar* d'un contorn és un recurs ben conegut i àmpliament utilitzat per a la descripció d'un contorn i el reconeixement de formes. La proposta realitzada en aquesta tesi és la seva aplicació en la descripció i reconeixement de *característiques locals* dels objectes, aplicant la transformació a petites regions de la imatge.

Aquest *reconeixement local* és aplicat al seguiment tridimensional dels objectes a partir del seguiment de les singularitats del seu contorn. Amb aquesta metodologia s'aconsegueix per una banda una alta reducció de la quantitat d'informació a processar, i una elevada robusteça davant l'oclusió que de forma esporàdica pugui patir alguna de les característiques locals seguides.

La proposta queda inclosa, per tant, en els sistemes de seguiment mitjançant visió per computador que es basen en un reconeixement local de la imatge. La metodologia de reconeixement proposada, però, no ha estat abordada abans en la bibliografia disponible.

A continuació s'enumeren de forma resumida les aportacions més rellevants del sistema de seguiment proposat:

- S'ha efectuat una optimització del sistema de reconeixement local utilitzat, amb el criteri cost-eficiència-temps per a una implementació *hardware* que permeti l'operació en temps real. S'ha determinat l'àrea òptima de transformació i el nombre necessari de mesures radials. Donada la baixa resolució de la codificació polar proposada, l'estudi de la seva eficiència s'ha completat amb un anàlisi de la incidència dels errors presents a la discretització i localització dels contorns.

- A partir d'aquesta transformació polar discreta de la imatge, s'ha desenvolupat un mètode eficaç per a la detecció i localització automàtica de les característiques locals dels objectes continguts a l'escena. Aquest mètode permet a més, fer una valoració d'aquestes característiques locals, de forma que a partir d'aquesta informació poden ser seleccionades de forma automàtica, el que permet una inicialització automàtica del seguiment.

- S'introdueix un sistema de validació del seguiment de les característiques locals a partir d'un anàlisi de coherència en les dades relatives a les seves trajectòries tridimensionals. Això permet que el seguiment tridimensional sigui assolit a partir d'un aparellament estèreo inicial i un seguiment posterior de la característica local en les dues imatges binoculars.

- S'ha implementat un processador específic d'imatge, de baix cost, que realitza la adquisició, extracció de contorns i codificació polar de la imatge en temps real (*video rate)*. Connectat a un PC estàndard, permet seguir tridimensionalment fins a quatre característiques locals a 50Hz. Per un nombre més elevat de característiques aconsegueix temps de processat 10 vegades menors que amb l'ús d'una placa d'adquisició d'imatges convencional.

INTRODUCCIÓ

0. Introducció

> *'The fact is that civilization requires slaves. The Greeks were quite right there. Unless there are slaves to do the ugly, horrible, uninteresting work, culture and contemplation become almost impossible. Human slavery is wrong, insecure, and demoralizing. On mechanical slavery, on the slavery of the machine, the future of the world depends.'*
>
> *OSCAR WILDE, 1891*

Els científics, de qualsevol àrea, tenen sovint la impressió d'estar aprenent molt d'ells mateixos a través de la seva feina. En el cas de la Robòtica aquesta apreciació és gairebé obligada. La connexió entre el camp d'estudi i nosaltres mateixos és inusualment elevada. A més a més, a diferència d'altres ciències que es decanten majoritàriament per l'anàlisi, en el cas de la Robòtica, amb una forta implicació tecnològica, la tendència és cap a la síntesi.

La Robòtica tracta en essència el desig de sintetitzar alguns aspectes de la funcionalitat humana fent servir la tecnologia disponible (mecanismes, actuadors, sensors, computadores...). Aquest desig de síntesi d'un mecanisme amb funcions "semblants" a l'home és tant antic com la pròpia història [Geduld,78]. La mitologia grega és plena d'exemples on alguns homes amb l'ajuda dels deus construïen tota sèrie d'artefactes (a vegades per pura diversió com en el cas del deu Hefaest).

La consecució d'aquest objectiu implica avui dia a moltes àrees d'investigació clàssiques, com ara, el Control, la Intel·ligència Artificial i la Visió per Computador entre d'altres. Pot ser per aquesta raó, la Robòtica ens fascina a molts de nosaltres.

La Robòtica pot ser definida com la connexió intel·ligent entre la percepció i l'acció [Brady,85]. Dins dels sistemes existents de percepció, la Visió per Computador és una de les fonts d'informació amb més potencial per a la Robòtica. És erroni, però, considerar la Visió per Computador i el Control, com a meres interfícies (I/O) de la Intel·ligència Artificial (IA).

Dins de la Visió per Computador es plantegen problemes d'aparença trivial però de difícil solució. Les eines utilitzades per atacar-los poden arribar a ser molt complexes. Moltes vegades la solució d'aquests problemes requereixen d'un tractament previ del senyal que subministra el sensor (la camera) o manipulacions matemàtiques sofisticades que disten molt del tractament simbòlic d'informació que acostuma a fer-se amb la utilització de tècniques de IA.

A més a més, quan la Visió per Computador és integrada en un sistema robòtic, com un mòdul més de percepció, el temps de procés juga un paper fonamental en l'avaluació de les solucions proposades. En aquest cas, cal que la resposta del sistema sigui en "temps real" , altrament les solucions poden ser correctes però no ser *útils*.

Estem en un moment en que la tecnologia sembla posar al nostre abast la possibilitat de donar solució a molts dels problemes plantejats dins de la Robòtica. No podem restar passius davant d'aquest repte. Aquesta tesi pretén donar una possible eina per a la solució d'un d'aquests problemes d'aspecte innocent i trivial que apareix dins de la Visió per Computador. Es tracta del seguiment d'objectes en un espai tridimensional mitjançant la estereovisió.

Al món de la biologia són prou coneguts la quantitat de mecanismes de seguiment que la natura ha desenvolupat. Són molts els sistemes biològics que han evolucionat per fer front a un món que es presenta molt canviant. Es poden trobar des de sistemes d'adaptació lenta com el seguiment de la llum que fan les plantes, fins als mecanismes de resposta més ràpida que permeten els depredadors aconseguir les seves presses.

Resulta evident la rellevància que la detecció de moviment té en el món animal. Aquesta qualitat no no és única dels vertebrats superiors (pensem per exemple en una mosca que esquiva la nostra mà) la qual cosa fa suposar que la detecció de moviment ha de consistir en quelcom *'hardware'* que apareix fins i tot en els éssers menys desenvolupats [Ullman,81]. No n'hi ha prou però amb la pura detecció, en molts casos cal, a més a més, localitzar tridimensionalment l'objecte que es mou de forma precisa, i en l'últim terme identificar-ho. No cal insistir en la importància que té el moviment en la *interpretació* de l'entorn.

La Visió per Computador ha evolucionat de la mateixa forma. Al seu inici es va ocupar d'extreure informació a partir d'imatges estàtiques. L' habilitat de fer front al moviment dels objectes, els canvis de forma, canvis d'il·luminació o canvis de punt de vista (projecció) és per moltes aplicacions essencial; indispensable per a totes les relacionades amb la Robòtica.

La implementació de mètodes útils per a l'anàlisi de seqüències d'imatges, a fi d'extreure informació relativa a aquests possibles canvis en la escena, té una doble dificultat: la pròpia de qualsevol procés d'anàlisi d'imatges, més la deguda a la necessitat d'una implementació que ha de donar resultats en temps real. Altrament el sistema no serà sensible als canvis. La dificultat augmenta encara més al considerar imatges estèreo.

Els sistemes de visió per computador assumeixen habitualment que els canvis produïts a la imatge són deguts al moviment de la camera i/o el moviment dels objectes que apareixen a la escena. Els objectes són sempre considerats rígids o quasi-rígids. Altres canvis no són considerats per la majoria d'autors [Jain,95]. Apareixen llavors quatre possibilitats de classificació dels sistemes de visió en funció de la naturalesa dels canvis:

1. Camera estàtica i objectes estàtics (SCSO).
2. Camera estàtica i objectes mòbils (SCMO).
3. Camera mòbil i objectes estàtics (MCSO).
4. Camera mòbil i objectes mòbils (MCMO).

El primer cas consisteix en l'anàlisi d'escenes estàtiques. Una gran part dels sistemes de visió pertanyen a aquesta categoria. Dins dels altres tres casos que inclouen algun tipus de moviment, el segon és amb molt el més estudiat i en el que s'han dedicat més esforços en els últims anys amb bons resultats. Moltes de les tècniques usades quan la camera és estàtica no són aplicables, però, quan la camera també es mou. Els últims dos casos tracten aquest problema i són especialment importants en aplicacions de robòtica mòbil. El cas més general és l'últim i és també possiblement el de més dificultat, motiu pel qual és el menys desenvolupat dins de la Visió per Computador.

En el treball que s'ha realitzat es presenta un sistema de seguiment automàtic d'objectes que pretén ser aplicable a qualsevol dels supòsits anteriors. El sistema ofereix a la sortida les trajectòries tridimensionals dels objectes seguits.

La memòria d'aquesta tesi s'ha estructurat en sis capítols que tracten, respectivament, dels punts següents:

En el primer capítol, i després de fer una breu introducció al problema de la percepció tridimensional, es descriuen els diferents mètodes existents. Es comença pels sistemes actius per arribar fins als sistemes passius i més concretament als basats en l'estereovisió.

Al segon capítol es revisen les diferents aproximacions al problema del seguiment d'objectius que han sigut portats a terme fins ara. Es dona especial èmfasi a la descripció dels enfocaments anomenats heurístics, que treballen amb seqüències curtes d'imatges i que permeten una resposta en *temps real*.

El tercer capítol es dedica a la descripció del sistema de seguiment tridimensional proposat. A cada apartat es detallen els diferents mòduls de processat de dades que composen el sistema. Així mateix, es presenten resultats del processament de dades realitzat per cada mòdul.

Al capítol quart es presenta la implementació *hardware* realitzada, la qual permet una elevada velocitat d'operació (pot arribar als 20 ms) del sistema de seguiment exposat al capítol anterior, i es presenten els resultats obtinguts pel sistema.

Al cinquè capítol s'analitzen les fonts d'error en la localització dels objectes que afecten al sistema de seguiment proposat. S'utilitzen tècniques estadístiques per afitar els errors que es produeixen en els diferents mòduls de processat de dades que integren el sistema, deguts bàsicament a la discretització de la imatge i a la localització dels contorns.

Al darrer capítol es presenten les principals conclusions i les aportacions realitzades, sempre en relació amb els resultats obtinguts. Finalment es contemplen algunes possibles aplicacions del sistema i s'apunten algunes línies de recerca a curt termini que s'han plantejat a partir del treball realitzat.

CAPÍTOL 1

1.Percepció tridimensional: antecedents i estat de l'art.

1.1.Introducció

Per arribar a la interpretació d'una escena, és molt convenient disposar d'una certa informació tridimensional sobre una imatge bidimensional. Tot i que hi han molts sistemes en procés de desenvolupament per realitzar aquesta mesura de distància dels punts continguts a l'escena, encara no hi han sistemes disponibles que permetin l'obtenció en temps real d'un mapa dens de profunditats a un cost competitiu per a la seva aplicació industrial.

Els mètodes d'adquisició d'aquests mapes de profunditat poden ser classificats de la següent manera:

1) **actius:** basats en l'emissió d'una radiació (llum, so, microones, ...) i en la mesura:

a)per temps de vol: comptant el temps que tarda en detectar-se la reflexió.

b)per triangulació: observant la direcció dels raigs reflectits.

2) **passius:** basats en la triangulació de punts homòlegs de múltiples imatges.

a)Estereovisió: imatges adquirides alhora des de dues, tres o més cameres fixes.

b)"Stereo motion": imatges d'objectes fixos adquirides en temps diferents des d'una mateixa camera en moviment [Jiang, 95].

c)"Structure from motion": imatges d'objectes en moviment adquirides en temps diferents des d'una camera fixa.

Els mètodes actius de temps de vol que utilitzen ultrasons, tenen l'avantatge de ser relativament econòmics i poc aparatosos, però normalment la utilització dels ultrasons no és suficient per adquirir un mapa de profunditats prou dens i acurat. A més a més presenta el problema de les reflexions especulars que dificulten la detecció dels objectes. Per contra, els mètodes de temps de vol amb llum làser requereixen de costosos aparells per fer les mesures, però la fan de forma molt precisa. Aquests mètodes són emprats en alguns prototipus de vehicles autònoms per a l'evitació d'obstacles.

Els mètodes de triangulació activa utilitzen un projector de llum estructurada que incideix en els objectes de l'escena, i una camera detecta les imatges obtingudes des d'una altra direcció [Shirai, 71]. Aquests mètodes no només s'han fet servir per la investigació, si no que també han estat utilitzats a la indústria, en aplicacions d'inspecció automatitzada i soldadura [Shirai, 92].

L'avantatge dels mètodes actius és l'obtenció de dades de profunditat sense excessius problemes de computació. Aquests mètodes però, només funcionen en casos limitats, allà on la normal a la superfície dels objectes no forma un angle massa elevat amb la direcció de la llum (o els ultrasons), de forma que es pot produir el reflex. A més els objectes han d'estar suficientment a prop per que els raigs reflectits siguin visibles.

En canvi, la percepció tridimensional passiva basada en la informació estèreo pot funcionar en una alta varietat de condicions, en principi, en totes aquelles en les quals la visió binocular humana funciona.

1.2.Triangulació activa

Molts dels treballs previs en adquisició de mapes de profunditat utilitzen el mètode de la triangulació activa amb llum estructurada [Agin,76][Oshima,83]. Un dels problemes clàssics d'aquest mètode era la seva lentitud, ja que cada imatge amb una franja de llum era adquirida cada 1/25 segons i per tant es necessitaven 10 segons per obtenir imatges amb 250 franges.

En lloc d'escombrar l'escena amb una sola franja de llum, es poden utilitzar múltiples plans de llum. El problema llavors és la necessitat d'identificar cadascuna de les franges de llum. Per fer-ho s'utilitzen diferents patrons codificats de plans verticals de llum i s'adquireix una imatge per cada patró. Amb aquest mètode només es necessiten $\log(n)$ imatges, sent n el número de franges, per obtenir la mateixa resolució [Minou,81]. Un sistema pràctic va ser desenvolupat per Sato, [Sato,87], que genera els patrons amb una màscara de cristall líquid. De totes formes aquest mètode és encara lent per poder seguir objectes en moviment.

Araki, [Araki,87][Araki,88], va proposar un sistema molt més ràpid que emprava una matriu de fototransistors en lloc d'una camera de televisió. Mentre un làser escombra l'escena, cada fototransistor registra el temps que tarda en arribar el pic d'intensitat de llum. Com que la direcció del raig de llum pot ser determinada a partir d'aquest temps, la posició de la franja de llum es pot calcular per triangulació. El sistema pot ser ràpid en principi perquè el temps de mostreig no està limitat a 1/30 segons. Amb el seu primer sistema, Araki obtenia 400 línies per segon amb una matriu de 47x47 fototransistors.

Altres sistemes també utilitzen un raig làser i dispositius sensibles a la posició (PSD) com l'exposat abans, bé bidimensionals [Kanade,81] o bé unidimensionals [Rioux,83].

Més tard Araki, [Araki,90][Araki,91], canvià la matriu de fototransistors per una matriu lineal de PSDs. Cada PSD detecta la posició del punt més brillant cada 5μsg. El primer sistema utilitzava 30 PSDs, i obtenia 30 imatges per segon amb una resolució de 30x128. Actualment, Araki treballa en un nou sistema que adquireix 100 imatges per segon amb una resolució de 128x128. Un dispositiu semblant està sent desenvolupat a la Universitat d'Osaka [Kida,88].

A la Universitat de Carnegie Mellon, el Dr. Kanade, [Kanade,93], ha implementat un sensor semblant de 32x32 fotoreceptors integrats en un sol xip VLSI, [Carley,90][Grauss,92], que permet obtenir fins a 250 imatges per segon amb una precisió de 0.5 mil·límetres. Això permet la utilització d'aquest sensor en tasques de seguiment tridimensional d'objectes [Simon,94] [Kanade,97].

A Europa també es treballa en aquesta línia, amb sistemes que donen resultats prou ràpidament com per fer seguiment tridimensional d'objectes [Lindsey, 95].

Contrastant amb aquests resultats extraordinaris que ofereix la triangulació activa, destaca l'escassa distància en la qual aquests mètodes són operatius amb precisió, del ordre de 300 a 1000 mm i els problemes d'ombres en la imatge de profunditats que generen els propis objectes de l'escena.

1.3. Estereovisió

La triangulació passiva de dues imatges pot ser realitzada mitjançant la visió binocular, anomenada també visió estèreo o estereovisió. Encara que la visió estèreo té un poder potencial alt, encara no és usada habitualment a la indústria. La raó principal és l'absència de procediments abordables pel processament d'imatge necessari, degut al cost computacional, que obliga a la utilització de supercomputadors i/o processadors específics d'imatge d'alt cost econòmic.

El problema fonamental en estereovisió és l'aparellament ("matching") dels punts homòlegs de les dues imatges. Una vegada la correspondència ha estat determinada, la posició tridimensional d'aquests punts és fàcilment calculada a partir de la disparitat (diferència en la posició d'un punt en les dues imatges) per triangulació [Binford, 83] [Cochran, 92].

Han sigut proposats molts algoritmes orientats a solucionar el problema de l'aparellament, els quals es poden classificar en dues categories:

a)basats en àrea (correlació local).

b)basats en característiques.

Els mètodes basats en àrea troben els punts homòlegs buscant el grau de similitud entre diferents parts (finestres) de la imatge esquerra i dreta. Els mètodes basats en característiques aparellen punts singulars d'una imatge amb punts singulars de l'altra, considerats homòlegs [Weng, 92] [Ahuja, 93].

Aquests mètodes es detallen a continuació, així com diferents estratègies per resoldre els problemes d'ambigüitat en la correspondència ja sigui entre finestres o punts singulars.

1.3.1. Estereovisió basada en àrea

La majoria d'aquests mètodes seleccionen alguns punts característics d'una imatge (o bé utilitzen tots els punts de la imatge) així com la subimatge que els envolta, i fan us de la restricció epipolar, és a dir, el seu homòleg descansa sobre la línia epipolar en la altre imatge. Llavors, una subimatge semblant es adquirida al llarg d'aquesta línia en l'altra imatge i es calcula la seva similitud amb la finestra original [Völpel, 95].

Hi ha dos problemes a l'hora de trobar la correspondència, la solució dels quals distingeix els diferents mètodes:

a)Com avaluar la similitud.

b)Com determinar amb precisió el punt homòleg a partir de la funció similitud.

Pel primer problema, una solució molt utilitzada és calcular la correlació entre les dues finestres. Això requereix un esforç computacional molt elevat ja que la correlació és la covariància de les dues imatges dividida per l'arrel quadrada del producte de la variància de les dues imatges. Normalment es redueix el càlcul dividint només per la variància d'una de les imatges.

Amb la intenció de reduir el temps de càlcul d'una forma més dràstica, es calcula la diferència entre les imatges, entenent aquesta com la suma del valor absolut de la diferència entre el valor dels píxels d'ambdues finestres ("gray-level matching"). Aquest mètode funciona bé si les dues cameres tenen les mateixes característiques, donen la mateixa intensitat de llum per pixel i la llum reflectida pels objectes no canvia en funció de la direcció d'observació, condicions totes dues difícils d'assegurar [Horn,86].

El següent problema presentat és que la correspondència entre punts homòlegs pot no ser biunívoca, ja que alguns punts de la imatge no tenen homòlegs a l'altra imatge, degut a que poden caure fora de la imatge o poden quedar ocults degut a la perspectiva [Geiger, 92]. Això implica que les correspondències no poden ser localitzades buscant el màxim de la funció similitud o el mínim de la diferència. Algunes propostes per solucionar aquest problema son la utilització de múltiples llindars [Yasuye,73] o l'anàlisi de la funció correlació al voltant del màxim [Nishihara,84].

El principal avantatge d'aquests mètodes basats en area és la seva facilitat d'implementació amb un preprocessador vectorial (SIMD), de cara a millorar el alt cost de càlcul implicat en les operacions. Aquests mètodes fallen però, si el nivell de gris és uniforme dins d'una regió de la imatge. De forma evident només donen bons resultats de localització en presència de contorns o gradients en la imatge. Llavors sembla més convenient realitzar de forma previa una extracció de característiques (globals o locals) presents en la parella d'imatges.

1.3.2 Estereovisió basada en característiques

Moltes i diferents característiques han estat seleccionades com a base per a la comparació entre imatges. Hom pot separar aquestes característiques en dues categories:

a)característiques globals.

b)característiques locals.

Les característiques globals, com ara la longitud i angle de segments rectes llargs o l'àrea i orientació d'objectes, poden ser aparellades més fàcilment. Aquest tipus de característiques són adequades per escenes amb objectes artificials i la seva detecció no és en general fàcil. Encara més, la selecció de les característiques globals adequades per fer l'aparellament depèn de l'escena que volem mesurar, com pot ser comprovat en la gran varietat de característiques proposades [Chongstitvatana, 92] [Sagüés, 92] [Tseng, 92] [Devy, 92] [Zhang, 92] [Lee, 94] [Zhang, 95].

Generalment, les característiques locals són més fàcils de detectar però pel contrari són més difícils d'aparellar, ja que en principi pot haver molts candidats semblants. Els mètodes de detecció de les característiques locals solen ser abordables des del punt de vista de la seva implementació *hardware* [Ohta, 87] [Ferrari, 92] [Courtney, 92] [Takeno, 92], mentre que la obtenció de característiques globals sòl ser mes costosa i difícil d'implementar amb un *hardware* específic.

Sembla ser que la visió animal detecta característiques locals per fer l'estereovisió [Mayhew, 81] [Ludwing, 94]. Així es desprèn d'experiències efectuades amb l'aparellament d'estereogrames de punts aleatoris que demostren que el problema de la correspondència és resol abans del reconeixement [Julesz,71].

La característica local més usada són els contorns, que normalment es detecten com el pas per zero de la imatge filtrada amb un filtre Gaussià [Marr,82]. Habitualment els contorns de baix contrast no són tinguts en compte o són guardats per una etapa posterior, reduint d'aquesta forma el número de punts característics candidats a l'aparellament [Nakayama,92]. La resolució d'aquest mètode depèn de la variància del filtre Gaussià. L'aparellament entre contorns es determina normalment amb un llindar sobre la diferència de les seves direccions [Olsen,90].

Aquesta restricció no és suficient per determinar de forma única la correspondència i ens hem de fer servir d'alguns heurístics. Per exemple, si les superfícies dels objectes són llises, un aparellament jeràrquic és efectiu si comencem aparellant característiques globals i gradualment obtenim correspondències de més resolució amb els contorns, tot tenint en compte les anteriors [Marr,82]. Com que les escenes "reals" consisteixen bàsicament en superfícies llises localitzades, un mètode de relaxació [Barnard,80] i un histograma de disparitats poden ser utilitzats per trobar distàncies semblants en un entorn local [Nishimoto,86].

També es pot utilitzar un difuminat jeràrquic i començar l'aparellament amb contorns presents entre àrees de l'escena amb gran diferència de contrast independenment de la seva "textura" interna [Horn,86].

1.3.3 Mètodes per avaluar l'aparellament

Ja s'han comentat anteriorment els problemes inherents a l'hora de solucionar el problema de la correspondència, tant en els mètodes basats en àrea com en els basats en característiques. Donada aquesta situació es fa necessari l'estudi de tècniques que permetin avaluar si els aparellaments entre imatges són correctes i augmentar així la fiabilitat del sistema. Una de les solucions més emprades pels diferents autors és l'ús d'aparellaments múltiples. Alguns exemples són:

a)test esquerra-dreta-dreta-esquerra.

Es tracta de realitzar cada aparellament dues vegades, una de l'esquerra a la dreta i a l'inrevés. Si la distància entre el punt inicial i l'homòleg trobat en l'aparellament final dreta-esquerra es prou petita, la correspondència és considerada vàlida [Hannah,89] [Fua,91] [Lassere, 95].

b)comparació entre múltiples estimacions de profunditat.

Donades N cameres (N>3), agafar una d'elles com a pivot i per a cada punt trobar els N-1 homòlegs en les altres imatges. Si les estimacions de profunditat són semblants l'aparellament es considera vàlid [Pietikäinem, 86] [Yoshida,92].

c)sumar múltiples superfícies de correlació.

Aquest mètode és semblant a l'anterior, només que en lloc de comparar les N-1 estimacions, suma les N-1 correlacions entre imatges i busca el millor aparellament "global". Per sumar els resultats, les disparitats obtingudes són normalitzades a una representació comú. Les diferents imatges poden ser adquirides movent una camera al llarg d'una barra [Moravec,79] o bé disposant d'una bateria de cameres [Kanade,92] [Okutomi,93]. Una proposta d'un maquinari específic per realitzar aquest aparellament en temps real es proposada en [Kanade,93].

d)test de la línia epipolar.

Donades tres cameres calibrades, trobar la correspondència dels punts de la imatge 1 en la 2 i de la imatge 2 en la 3. Si la distància entre el punt trobat a la imatge 3 i la línia epipolar corresponent al punt de la imatge 1 és petita, l'aparellament és vàlid [Ayache,87].

e)test de transitivitat.

Semblant a l'anterior, però amb l'avantatge que les cameres no han d'estar calibrades. Aquí, donades tres cameres, trobar la correspondència dels punts de la imatge 1 en la

2, de la imatge 2 en la 3 i de la imatge 3 en la 1. Si la distància entre el punt inicial en la imatge 1 i l'aparellament final és petita, la correspondència és vàlida [Bolles,93].

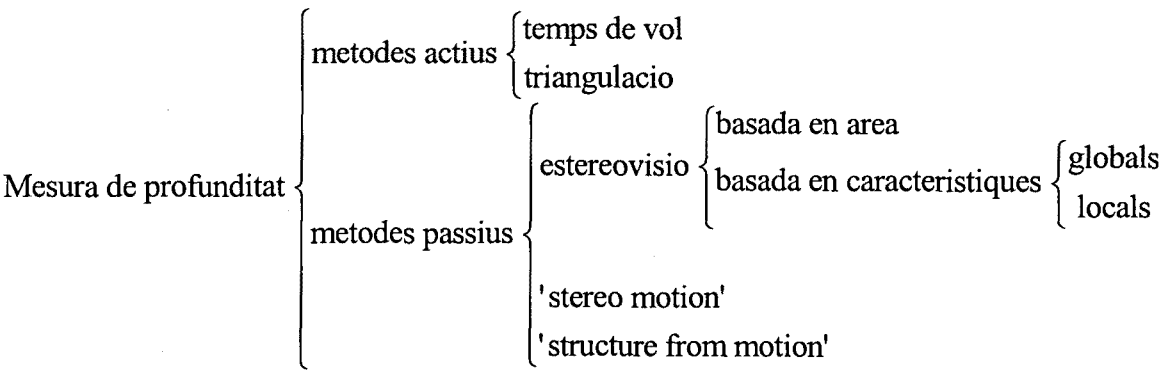
f)test de moviment relatiu del vehicle.

Donades dues imatges estèreo adquirides al llarg del temps des d'un vehicle amb moviment conegut o fàcilment calculable, obtenir la posició (x,y,z) dels punts de la imatge per a un instant i tornar a calcular la posició (x,y,z) d'aquest mateixos punts posteriorment. Si la posició ha variat allò que estava previst degut al moviment del vehicle, l'aparellament és correcte, si no pot ser o bé incorrecte, o bé que els punts seleccionats pertanyen a un objecte en moviment [Bolles,93]. Aquest sistema requereix fer un seguiment ("tracking") dels punts dins de la seqüència d'imatges.

g)test de moviment relatiu dels objectes.

Si el moviment del vehicle no és conegut, o els objectes es mouen dins l'escena, llavors es pot seleccionar un punt arbitrari com a referència i calcular les posicions (x,y,z) dels altres punts relatius a aquest. Si la seva distància relativa és estable entre imatges consecutives l'aparellament és considerat vàlid [Moezzi,91]. Aquest sistema també necessita d'un seguiment dels punts dins de la seqüència d'imatges.

1.4. Esquema dels mètodes de percepció tridimensional



CAPÍTOL 2

2. El problema del seguiment automàtic d'objectius mitjançant visió per computador.

2.1. Introducció

Dins dels sistemes de seguiment automàtic d'objectius hom ha de distingir en una primera aproximació entre els sistemes actius i els passius.

Els sistemes actius com ara el radar, els ultrasons i el làser proporcionen de forma explicita mapes de profunditat que poden ser molt útils i fàcils de tractar per les aplicacions de seguiment d'objectius. Aquests sistemes, però, a més de ser comparativament molt costosos, utilitzen sistemes d'escombrat mecànics de poca fiabilitat i tenen un consum excessiu, la qual cosa els fa impracticables en moltes aplicacions.

És per això que la captació electromagnètica passiva, i la visió per ordinador en particular, constitueix actualment una alternativa atractiva als sistemes actius, donant unes resolucions espacials i unes freqüències de mostreig que no poden ser superades pels sistemes actius i que són determinants per algunes aplicacions.

La investigació realitzada en els últims vint anys en el camp del seguiment d'objectius mitjançant la visió per ordinador ha aportat bones solucions a un gran nombre de problemes en tota mena d'aplicacions com ara el seguiment d'avions i míssils, aplicacions robòtiques [Amat,92], la metereologia, el control del trànsit [Aranda,93], estudis de comportament i psicologia [Aranda,94], etc.

2.2. Definició del problema

El seguiment d'objectius actiu o passiu es defineix com el processament de mesures obtingudes d'un objectiu ('target') amb la intenció de mantenir una estimació del seu estat actual [Bar-Shalom,95].

L'estat d'un objectiu consisteix típicament de:

- Les seves components cinemàtiques (posició, velocitat, acceleració, celeritat)
- Una descripció de l'objectiu (longitud d'ona irradiada, característiques espectrals, grandària, forma, etc.)
- Uns paràmetres constants o de variació lenta (derives tèrmiques, velocitat de propagació, etc.)

En concret, pel cas del seguiment mitjançant visió per computador, la representació més utilitzada de l'estat combina característiques cinemàtiques més una certa descripció, com per exemple la seva posició i la seva àrea.

Les mesures d'un objectiu són les observacions, alterades pel soroll, de l'estat de l'objectiu, que es poden obtenir a partir de:

- Estimació directa de la posició.
- Distància i/o angle respecte el sensor.
- Longitud d'ona (en cas de visió: color).
- Diferència de temps d'arribada del senyal entre dos o més sensors.
- Diferència de freqüències observades entre dos sensors (degut al desplaçament Doppler).

En el cas de la visió per ordinador les mesures corresponen als tres primers tipus.

Les mesures d'interès en la majoria d'aplicacions de seguiment no són dades directes proporcionades per un sensor, si no sortides d'un sistema de detecció i processament del senyal, tal i com mostra el següent esquema:

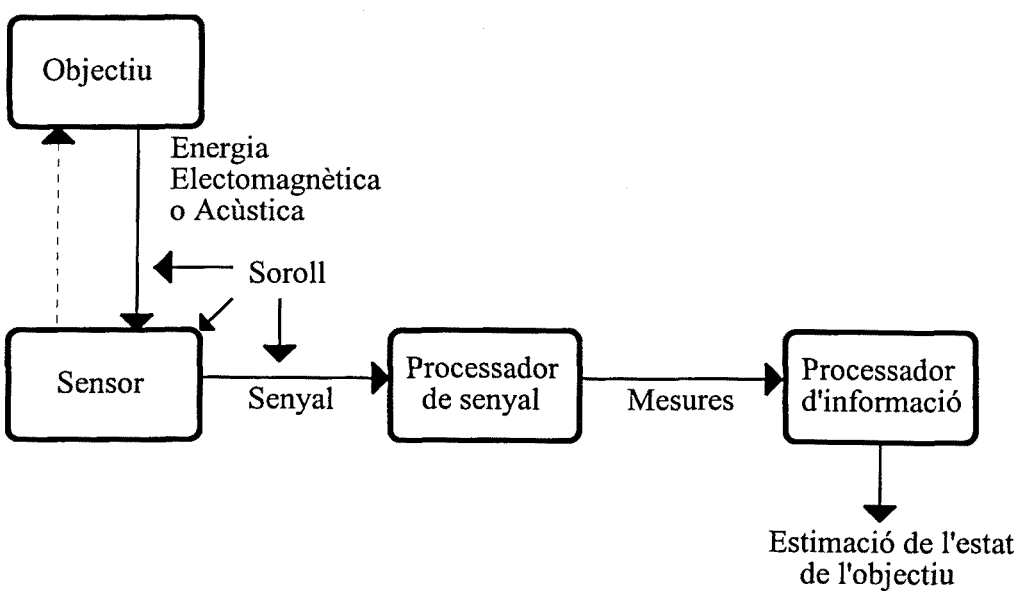


Figura 2.1. Esquema general d'un sistema de seguiment

A partir d'un conjunt de mesures que han estat associades amb un mateix objectiu, es pot fer una estimació de l'estat de la seva trajectòria. Aquesta estimació de la trajectòria és anomenada técnicament *'track'*.

Els principals problemes del seguiment d'objectius són:

- La detecció de l'objectiu.
- La definició de l'estat inicial de l'objectiu.
- La incertesa de les mesures.
- El procés d'associació de dades a partir de les mesures obtingudes.

Les mesures obtingudes de l'estat de l'objectiu a través del processador de senyal no es corresponen amb tota fidelitat amb l'estat real degut al **soroll**. Aquest soroll pot presentar-se com:

- Pertorbacions de l'energia que l'objecte irradia (o reflexa) degudes a variacions en el medi de propagació (temperatura, humitat ...).
- No linealitats en la sensibilitat del transductor respecte al valor de l'energia rebuda.
- Soroll electromagnètic sobre el senyal proporcionat pel sensor.
- Imprecissió en la conversió analògic/digital deguts al mostreig i la discretització de la codificació.

Per poder seguir un objectiu és necessari establir les **condicions inicials** del seu estat, a partir de les quals podrà començar el procés d'associació amb les noves mesures. La solució habitual per a aquest problema consisteix en donar per conegut l'estat inicial de l'objectiu. Això vol dir tenir un coneixement previ de l'objectiu a seguir, tant de les seves components cinemàtiques, com de les característiques que el defineixen. Pocs autors dediquen esforços en aconseguir mètodes de detecció i localització automàtica dels objectius a seguir, sense conèixer almenys de forma parcial l'estat inicial dels objectius [Rao, 93].

El principal problema del seguiment d'objectius, encara més greu quan es pretenen seguir múltiples objectius, radica en el **procés d'associació o correlació de dades** relatives als objectius a partir de les mesures obtingudes, l'origen de les quals, no hem d'oblidar, conté un cert nivell d'incertesa.

En l'associació de dades relatives al seguiment de múltiples objectius hem de distingir tres tipus:

- Associar una mesura a una mesura (inicialització d'un *'track'*).
- Associar una mesura a un *'track'* (manteniment d'un *'track'*, seguiment).
- Associar un *'track'* a un altre *'track'* (fusió de *'tracks'*).

En el següent apartat s'abordarà el problema de la inicialització i del seguiment. El problema de la fusió de *'tracks'* pertany al problema genèric d'associació de conjunts de dades, problema que no serà tractat en aquesta tesi.

2.3. El problema de l'associació de dades

Tal com s'ha esmentat anteriorment, la finalitat del seguiment és l'estimació de l'estat actual de l'objectiu, a partir del processament de les mesures obtingudes.

Bàsicament, existeixen dos tipus de models d'associació de dades a partir dels quals es dissenyen els algorismes de correlació de dades:

- models probabilistes.
- models deterministes.

En els **models probabilistes**, s'utilitzen taules baiesianes en les quals les probabilitats d'aconteixements individuals (per exemple, una mesura X pertany a l'objectiu A) són processades i utilitzades en els algorismes d'estimació de l'estat actual. El paràmetre que es pretén estimar es considera una variable aleatòria, la qual porta associada una funció densitat de probabilitat [Sharp, 95].

En els **models deterministes** l'origen de la mesura és conegut i cert (o es assumit així per conveniència). El paràmetre a estimar té un únic valor possible obtingut amb alguna de les següents tècniques:

1. S'obté un conjunt d'associacions candidates, escollint la més probable, tot assumint el fet de que no té perquè ésser la correcta.
2. S'efectua el test d'hipòtesi clàssic, acceptant que l'associació està sotmesa a un error de probabilitat, però tractant-la com si fos certa.

Els resultats de l'associació determinista són habitualment utilitzats en els algorismes estàndard d'estimació d'estat, tal com el filtre de Kalman (lineal o estès) [Bar-Shalom,88].

Tant en el cas dels models probabilistes com en el dels models deterministes, tots els algorismes de seguiment utilitzen algun model del comportament dinàmic de l'objectiu. Encara que els models dinàmics varien molt en complexitat, generalment comparteixen l'assumpció de que el moviment de l'objectiu està governat per lleis físiques ben conegudes ja siguin velocitat limitada, acceleració limitada o nul·la (moviment uniforme), moviment circular, parabòlic, harmònic, etc.. Aquests models dinàmics han de ser construïts per tenir sempre present que els objectius poden ser pertorbats per altres causes externes no incloses en el model i que poden arribar a invalidar el comportament previst.

Aquests canvis imprevisibles en el moviment de l'objectiu representen el major repte en el disseny d'un sistema de seguiment, particularment quan es combinen amb incerteses en l'origen de la mesura. Els intents de superar aquesta dificultat amb els objectius que difícilment s'ajusten amb un model de comportament dinàmic, anomenats maniobrables (*'maneuvers targets'*), passen per una descripció més acurada del seu valor d'estat, ja sigui en les seves components cinemàtiques (inclusió de les n-èssimes derivades), i/o en la descripció de l'objectiu (enfortiment del **reconeixement**). Aquestes dues alternatives, que no són incompatibles, marquen una diferència elevada quan al tipus d'algorisme de seguiment i la seva metodologia, sent la primera clarament predictiva i la segona més determinista.

En el cas del seguiment d'objectius maniobrables, adquireix rellevant importància la disminució del temps entre mesures de l'estat d'un mateix objectiu, o dit d'una altra manera, l'augment de la **freqüència de mostreig** del nostre sistema de percepció. Augmentant la freqüència s'aconsegueix disminuir la variació de l'estat de l'objectiu i d'aquesta manera es simplifica el problema d'associació de dades entre mesures consecutives.

2.3.1 El problema tecnològic

Per aconseguir una freqüència de mostreig elevada, la primera restricció la imposa el sistema d'adquisició de les mesures, la seva freqüència de mostreig i la velocitat de transferència del senyal obtingut. Per tant, la freqüència màxima de mostreig del sistema de seguiment vindrà determinada pel tipus de sensor de que es disposa per fer l'adquisició.

Després de la seva adquisició, el senyal procedent del sensor ha de ser processat amb la finalitat d'extreure el valor de les mesures de l'estat dels objectius. Aquestes mesures seran utilitzades posteriorment en el procés d'associació de dades.

Hom creu important establir aquí una diferència important entre dos tipus de processat:

- Processat *off line.*
- Processat *on line.*

En el primer cas les mesures proporcionades pel sistema de percepció han de ser memoritzades pel seu processat posterior. La freqüència de mostreig depèn del sensor utilitzat, del temps de procés del senyal que tinguem, i de la velocitat d'adquisició de dades del dispositiu que hagi de memoritzar-les. El pitjor dels temps determinarà la freqüència. En el cas del seguiment mitjançant visió per ordinador, on el temps de procés del senyal és el més restrictiu, s'acostuma a memoritzar directament el senyal del sensor (de forma analògica o digital), deixant pel posterior procés *off line* tant el processat d'aquest senyal com l'associació de les dades que d'ell obtindrem. Això representa la utilització d'uns dispositius de grabació de gran volum d'enmagatzematge i alta velocitat d'accés (>3Mbytes/sg), és a dir, poc econòmics.

En el segon cas, processat *on line,* tot el processat s'ha de fer en **temps real**. Aquesta és l'única possibilitat de processat en aplicacions de control on el resultat del seguiment està directament lligat a l'acció del sistema. La freqüència de mostreig depèn aquí per tant, de la suma del temps d'adquisició, del temps de procés del senyal i del temps necessari per fer l'associació de les mesures obtingudes. Aquesta és la raó per la qual, quan el seguiment es fa *on line*, el procés del senyal i l'associació de dades han de ser realitzats en el temps més curt possible, sumant-se així al problema intrínsec de la correlació de dades, el problema de minimitzar el seu temps de càlcul (**cost computacional**).

Per minimitzar el temps de càlcul existeixen diferents propostes, entre les quals podem destacar:

- Reducció del vector d'estat (dificulta el problema de l'associació de dades).
- Utilització d'arquitectures multiprocessador estàndard que permetin accelerar el càlcul (problema del cost econòmic del maquinari i del programari necessari).
- Disseny de processadors específics basats en maquinari de cost reduït que permetin processar el senyal en temps real i oferir dades que facilitin la seva correlació mitjançant programari en un temps curt.

2.4. Estimació del moviment

Per realitzar l'estimació dels paràmetres que defineixen el moviment d'un objectiu s'han proposat dos enfocaments contraposats:

- Estimació a partir de tècniques de filtrat òptim.
- Estimació a partir de tècniques heurístiques.

En la primera proposta es tracta de modelitzar la trajectòria de l'objectiu que és objecte de seguiment mitjançant models estocàstics que permeten estimar els paràmetres a partir d'eines estadístiques . Un exemple de filtre que pertany a aquest enfocament àmpliament utilitzat per molts autors és el filtre de Kalman [Kalman,60][Kalman,61].

Aquests mètodes només resulten adients quan hi ha incertesa en l'aparellament d'un gran número de punts, és a dir, quan l'algorisme de reconeixement és inexistent o bé no aconsegueix discernir entre els possiblement múltiples objectius, i sempre que el model dinàmic dels objectius sigui conegut i prou correcte [Chang,89]. Es per això que aquests algorismes necessiten de seqüències llargues d'imatges, normalment vàries desenes, per poder convergir [Brown,88] [Yao, 95] [Weinshall, 95]. A més a més necessiten d'un model fidedigne del soroll, tant de l'objectiu com de les mesures, a fi de donar bones estimacions.

A més a més, i tal com apunten Lee i Lin [Lee,88], els objectius maniobrables, els més difícils de seguir, no són fàcilment modelitzats per cap tipus de procés senzill, i requereixen d'una complexitat computacional usualment molt elevada per aquest tipus de seguiment.

Els mètodes heurístics d'estimació de moviment, prescindeixen del filtrat òptim de la informació, i per tant ofereixen la possibilitat de realitzar el seguiment sense considerar una seqüència llarga d'imatges (en alguns casos només dues imatges son requerides), i per tant, poden treballar en temps real amb uns requeriments computacionals menors. Aquests mètodes sempre consideren una **associació de dades determinista**.

Dintre d'aquest enfocament heurístic existeixen tres aproximacions que són fonamentalment diferents a l'hora de fer el seguiment.

1. Seguiment basat en la **detecció de moviment** dins d'una escena.
2. Seguiment basat en la **segmentació** dels objectius.
3. Seguiment basat en el **reconeixement** dels objectius.

2.4.1 Seguiment basat en la detecció de moviment

Aquests mètodes basen el seguiment dels objectius en la localització de conjunts de píxels de l'escena que tenen una variació d'intensitat al llarg del temps. Si suposem que el valor dels píxels correspon sempre a la mateixa direcció en l'espai, o sigui si la camera es fixa, una variació en el valor entre imatges consecutives indicarà un canvi produït en aquella direcció i per tant un moviment.

Existeixen tres tècniques que treuen profit d'aquesta metodologia de detecció per estimar el moviment:

1. Imatge diferència.
2. Anàlisi del domini freqüencial.
3. Flux òptic.

2.4.1.1. Imatge diferència

La **imatge diferència** és la solució més senzilla a l'hora de detectar canvis entre imatges. Consisteix en l'obtenció d'una nova imatge binària resultat de realitzar una resta pixel a pixel entre una imatge patró i una nova imatge. Per a cada pixel, si la resta supera un cert llindar el resultat que se li assigna és 1, altrament és 0. Pertanyen a aquesta categoria publicacions tals com [Chen, 92] [Fukui, 92] [Rao, 93] [Welch, 93].

Suposem ara que la imatge patró conté només un fons estacionari i que la camera es manté fixa sobre la mateixa escena. Si ara comparem aquesta imatge patró amb una imatge posterior que conté un objecte mòbil, la imatge diferència eliminarà tots els objectes estacionaris, deixant a 1 tots els píxels que pertanyen a l'objecte mòbil (suposant que l'objecte té valors diferents al fons, ja que altrament seria impossible seguir-ho amb visió per ordinador).

Una restricció important que imposa aquest mètode és l'estabilitat de la il·luminació de l'escena per tal que els valor dels píxels del fons estacionari no es vegi alterat per sobre del llindar d'error admès en la comparació. Deguts als petits canvis d'il·luminació, al petit moviment que pugui tenir la camera i a l'error de discretització de la imatge, és habitual en aquest mètode l'aparició d'un contorn de píxels a 1 al voltant dels objectes estacionaris al fer la diferència amb el patró original. Aquest soroll obliga en la majoria dels casos a aplicar algun algorisme de filtrat selectiu per eliminar aquests píxels aïllats. També es poden eliminar aquests píxels sorollosos calculant la imatge de diferències acumulades [Jain,81][Jain,83][Gonzalez,87].

Si no podem garantir l'estabilitat de la il·luminació, però suposem que la variació de llum és petita entre imatges, podem realitzar la diferència entre dues imatges consecutives, de manera que els objectes estacionaris queden eliminats, mentre que els objectes en moviment presenten a la imatge diferència unes franges de píxels amb valor igual a 1, que indiquen la direcció i el sentit del moviment. Un altra solució

consisteix en realitzar una actualització dinàmica del patró a cada imatge, mitjançant un filtre recursiu, la qual cosa permet una major estabilitat en front de canvis sobtats d'il·luminació [Amat,84] [Aranda,94].

2.4.1.2. Anàlisi del domini freqüencial

Les tècniques basades en l'**anàlisi del domini freqüencial** realitzen la transformada de Fourier de la informació continguda en una seqüència d'imatges. Aquesta transformació s'utilitza amb l'objectiu d'identificar el moviment que conté una determinada freqüència al domini transformat [Rajala,83].

La informació bàsica necessària consisteix en la projecció de la imatge sobre els eixos x i y. Aquestes projeccions són ponderades per un factor exponencial que identificarà cada posició i es fa el sumatori d'elles a cada nova imatge (t) :

$$g_x(t,k_1) = \sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y,t)e^{j2\pi k_1 x\Delta t} \qquad t = 0,1,...,T-1$$

$$g_y(t,k_2) = \sum_{y=0}^{N-1}\sum_{x=0}^{M-1} f(x,y,t)e^{j2\pi k_2 x\Delta t} \qquad t = 0,1,...,T-1$$

Si ara realitzem les tranformades de Fourier unidimensionals d'aquestes expressions al llarg d'una seqüència d'imatges:

$$G_x(u_1,k_1) = \frac{1}{T}\sum_{t=0}^{T-1} g_x(t,k_1)e^{-j2\pi u_1 t/T} \qquad u_1=0,1,...,T-1$$

$$G_y(u_2,k_2) = \frac{1}{T}\sum_{t=0}^{T-1} g_y(t,k_2)e^{-j2\pi u_2 t/T} \qquad u_2=0,1,...,T-1$$

obtindrem una relació freqüència-velocitat donada per:

$u_1 = k_1 v_1$ a l'eix de les x

$u_2 = k_2 v_2$ a l'eix de les y

Un pic en les freqüències u_1 i u_2 ha de ser interpretat com un moviment d'algun dels objectes que hi ha a l'escena a velocitat (v_1,v_2).

El principal avantatge d'aquest mètode respecte a la diferència d'imatges és la seva alta inmunitat al soroll, ja que al llarg d'una seqüència d'imatges el aparent moviment degut al soroll aleatori queda transformat a una freqüència propera al zero, igual que tots els píxels del fons.

Com a inconvenients estan l'elevat cost computacional que representa la transformada

de Fourier, i la impossibilitat de localitzar l'objecte que provoca el moviment, ja que el resultat de l'anàlisi freqüencial només dona informació sobre la velocitat dels objectes que apareixien en l'escena. A més a més, si hi ha més d'un objecte en moviment resulta impossible associar els pics de freqüència en les x amb els de la y (diferents velocitats).

2.4.1.3. Flux òptic

S'anomena flux òptic al mapa de velocitats generat sobre el pla de la imatge degut al moviment aparent de la intensitat dels píxels de la imatge. Aquest moviment aparent por ser causat per:

- la projecció sobre la imatge dels objectes que es mouen en 3D
- el moviment de la camera respecte a l'escena
- una seqüència d'imatges que donen la il·lusió de moviment

També un moviment de la font d'il·luminació pot provocar una variació en el fluxe òptic sense que hi hagi moviment per part de la camera o dels objectes de l'escena.

La dificultat intrínseca del seguiment amb flux òptic radica en la obtenció del mapa de velocitats [Aloimonos, 91] [Li, 93]. Si tenim que *E(x,y,t)* és el nivell de gris de la imatge a la posició *(x,y)* a l'instant *t*, llavors si *u(x,y)* i *v(x,y)* són les components x i y del vector del flux òptic en aquest punt, llavors vol dir que:

$$E(x+u\delta t, y+v\delta t, t+\delta t) = E(x,y,t) \qquad \text{per un petit interval de temps } \delta t.$$

Ara podem expandir la part esquerra en la seva sèrie de Taylor:

$$E(x,y,t)+\delta x\frac{\partial E}{\partial x}+\delta y\frac{\partial E}{\partial y}+\delta t\frac{\partial E}{\partial t}+e=E(x,y,t)$$

On *e* conté els termes d'ordre més gran o igual a dos, que poden ser despreciats.

Si derivem ara aquesta expressió respecte al temps, ens dona l'equació coneguda com l'equació de restricció del flux òptic [Horn,86]:

$$\frac{\partial E}{\partial x}\frac{dx}{dt}+\frac{\partial E}{\partial y}\frac{dy}{dt}+\frac{\partial E}{\partial t}=0$$

o de forma abreviada

$$E_x u + E_y v + E_t = 0$$

Les derivades respecte a x, y i t són obtingudes directament a partir de la imatge E, i són habitualment aproximades de forma local. La dificultat a l'hora d'aplicar l'equació anterior radica en el fet que conté dues incògnites u i v. La equació, per tant, descriu una línia de possibles solucions on és impossible discriminar-ne una sense imposar alguna altra restricció com per exemple la "suavitat" (*smoothness*) en la velocitat dels objectes [Horn,81][Schunk,86][Duncan,88].

Com calcular el flux òptic de tota la imatge és actualment un problema computacionalment prohibitiu i no permet trobar solucions úniques (*ill-posed problem*), una alternativa popular ha estat calcular el flux òptic només en uns quants punts singulars de la imatge [Ballard,82] [Bhanu,90] [Murray, 94]. Resulta evident, a partir de l'anàlisi de la equació, que només pot ser calculat el flux òptic allà on el gradient de la imatge és diferent de zero. En el cas d'un contorn, només pot ser calculat en la direcció perpendicular al mateix, sobre la qual hi ha gradient.

També han hagut esforços per tal d'aconseguir calcular el flux òptic d'una imatge completa en temps real, però el cost del maquinari associat a les solucions proposades no sembla raonable. Un dels exemples més recents és el processador ISHTAR, desenvolupat per la Universitat d'Osaka conjuntament amb Fujitsu, el qual ha estat dissenyat específicament per a aquesta tasca (encara que com a processador programable pot servir per altres tasques) i el seu cost al mercat supera els 100.000$.

2.4.2. Seguiment basat en la segmentació

En determinades circumstàncies la segmentació d'una escena proporciona bastant informació per identificar els píxels que pertanyen a un objecte (o a una part d'aquest) de forma unívoca. El mètode requereix d'una funció booleana que aplicada sobre cada pixel de la imatge (i/o els seus veïns) diu si el pixel pertany o no a l'objecte. La segmentació, tot i haver de ser aplicada a la totalitat de la imatge, és utilitzada sovint per fer seguiment, donat que és un procés de cost computacional fàcilment optimitzable, ja que opera sobre la informació proporcionada pel sensor i te un caràcter fortament local. Aquesta localitat del processament fa que sigui idònia la seva implementació en màquines SIMD, o en preprocessadors de senyal fets a mida.

La classificació d'aquest mètode depèn de les tècniques de segmentació utilitzades, sent les més habituals:

- Binaritzat.
- Color.
- Contorns.
- Textures.

Una vegada segmentat l'objecte, la seva localització en el pla de la imatge pot ser trobada de forma immediata a partir de la selecció arbitrària d'alguna de les coordenades dels píxels que pertanyen a l'objecte. Una solució molt més robusta és trobar la posició de l'objecte a través d'un simple càlcul del **centroide**, definit com el moment de primer ordre (mitjana) de les coordenades dels píxels que pertanyen a l'objecte. Així mateix la seva orientació pot ser calculada, encara que només en determinats casos, a partir dels moments de segon ordre d'aquestes coordenades [Horn,86].

D'aquesta manera el centroide (X,Y) de l'objecte ve donat per les formules:

$$X = \frac{\sum x}{n} \qquad Y = \frac{\sum y}{n}$$

i l'orientació pot ser calculada a partir de:

$$a = \sum (x - X)^2$$

$$b = 2\sum (x - X)(y - Y)$$

$$c = \sum (y - Y)^2$$

de la següent forma:

$$\theta = \frac{1}{2} arctg(\frac{b}{a - c})$$

Existeixen algunes propostes per accelerar aquests càlculs, com l'ùs de projeccions de la imatge (*profiles*) [Horn, 86], l'aplicació de la transformada de Hadamard [Lai, 88] o l'aplicació del filtre de Hatamian [Hatamian, 86] [Li, 93]. [Venkateswarlu,95] va proposar un mètode ràpid pel càlcul de la mitjana i la variància (moments de primer i segon ordre). Recenment, [Martinez, 96] mitjançant tèniques matricials aconsegueix accelerar el càlcul de moments d'ordre fins i tot superior.

El seguiment de l'objecte s'aconsegueix a base de localitzar de forma successiva l'objecte segmentat a cada nova imatge. Si la segmentació de la imatge pot ser realitzada de forma ràpida i així mateix la localització, tindrem un sistema d'alta freqüència de mostreig, característica habitual en aquests sistemes.

Aquest mètode, però, imposa una sèrie de restriccions molt fortes a la segmentació, com són:

- que la detecció de l'objecte és sempre efectiva.
- que la segmentació ha seleccionat de forma correcta tots o suficients píxels que pertanyen a l'objecte (o a la part que volem seguir).

- que la segmentació ha seleccionat només aquests píxels i cap altre més.
- que els píxels seleccionats sempre es corresponen amb els mateixos punts de la superfície de l'objecte.

Aquestes restriccions, i especialment l'última, són difícils de complir per les tècniques de segmentació actuals degut principalment a problemes d'il·luminació i orientació relativa de l'objecte amb la camera. Es per això, que aquest mètode de seguiment resulta poc fiable, i només pot ser utilitzat per aplicacions en entorns que siguin adaptables.

Moltes vegades l'entorn de l'objectiu no pot ser adaptat de forma adequada per garantir una segmentació de l'objectiu de forma única, sinó que a més a més dels píxels que pertanyen a l'objectiu també es seleccionen altres parts de l'escena. Això pot ser degut a que el fons no sigui homogeni o bé a que hi ha més d'un objectiu amb les mateixes característiques. En aquests casos, i sempre que tinguem garantit que el contacte entre els píxels segmentats indica que pertanyen al mateix objecte, es sol aplicar una de les següents tècniques per separar els objectius:

- l'etiquetatge
- les finestres

Aquestes tècniques permeten un seguiment simultani de múltiples objectius basat en la segmentació.

L'etiquetatge ("labeling") és un procediment d'agrupació selectiva dels píxels que pertanyen a cadascuna de les diferents regions de la imatge que han estat seleccionades per la segmentació. L'agrupació es realitza en base a la conectivitat dels píxels de cada regió, normalment mirant els vuit píxels veïns. Per aquesta raó, els algorismes d'etiquetatge solen ser recursius, utilitzant la tècnica del creixement de regions ("region growing"), per simplificar la seva implementació. Els algorismes iteratius, que escombren la imatge de dalt a baix una sola vegada, poden ser més ràpids de cara a utilitzar-los per fer el seguiment de les diferents regions.

El seguiment dels diferents objectius pot ser aconseguit seguint els següents passos a cada nova imatge:

1. segmentació
2. etiquetatge
3. càlcul del centroide de cada regió.
4. associació de centroides per proximitat geomètrica.

Els processos d'etiquetatge i càlcul del centroide poden ser realitzats de forma paral·lela pel mateix algorisme, optimitzant el temps de execució i per tant augmentant la freqüència de mostreig del sistema de seguiment. Una forma de enfortir aquest mètode de seguiment és insertar un procés de reconeixement entre aquests dos processos. Aquesta opció és habitualment descartada en el cas del seguiment en temps real, ja que el procés de reconeixement baixa de forma dràstica la freqüència de mostreig.

El mètode de les **finestres** consisteix en aplicar la mateixa metodologia de segmentació i localització d'un únic objectiu dins d'una imatge, a cada objectiu dins d'una finestra (subimatge) que l'envolta. Aquesta finestra s'anomena usualment **finestra de seguiment** (*"tracking window"*). Una finestra de seguiment està caracteritzada per una posició (x,y) dins de la imatge i unes dimensions (Δx,Δy), o bé per dues posicions, (x_{min},y_{min}) i (x_{max},y_{max}), corresponents als vèrtexs superior-esquerra i inferior-dreta.

El procediment habitual consisteix en realitzar el segmentat de la imatge per tots els píxels sota el mateix criteri, i després calcular el centroide del conjunt de píxels segmentats que hi ha a cadascuna de les finestres de seguiment. O sigui, la finestra actua com una restricció geomètrica en la localització i velocitat de l'objecte, assumint que conté la totalitat de l'objecte i de forma única. S'assumeix també que la segmentació de l'objectiu és fiable dins de cada finestra de seguiment, i de fet el mètode permet utilitzar diferents tècniques de segmentació dins de cada finestra. Com que no s'han d'establir correspondències entre els centroides dels objectes seguits, aquest mètode és més ràpid que el de l'etiquetatge. Tot i així, apareixen tres problemes:

- Quina ha de ser la posició de cada finestra?
- Quines han de ser les seves dimensions?
- Que fer si dues finestres es solapen?

La posició i dimensions inicials de la/es finestra/es acostuma a ser donada al sistema de seguiment pel propi usuari. De vegades s'indica al sistema no la posició inicial de l'objectiu si no la posició (o posicions) en la qual l'objectiu es esperat que aparegui en escena.

Per obtenir la posició inicial de forma automàtica, cal tenir un coneixement previ de l'objecte a seguir i aplicar després de la segmentació de la primera imatge algun algorisme d'etiquetatge que separi les diferents regions (objectius) que poden aparèixer. Una altra solució és aplicar a l'inici del seguiment, un procés de reconeixement que identifiqui els objectius, tècnica que serà descrita en el següent apartat. La localització inicial de les finestres pot ser aconseguida, llavors, a través d'un càlcul del centroide de cada regió. Aquest procés haurà de ser repetit cada vegada que vulguem conèixer si hi ha presents nous objectius a l'escena (inicialització dels '*tracks*').

Quant a la dimensió de la finestra, resulta evident que ha d'incloure tota la regió segmentada, i per tant ha de tenir com a mínim els valors de la finestra que envolta la regió (valors mínims i màxims de les coordenades dels seus píxels). Però, com que s'espera que en la següent imatge l'objectiu aparegui desplaçat respecte a la posició anterior, les dimensions de la finestra de seguiment hauran de ser més grans que les de l'objectiu, amb la finalitat de garantir la seva inclusió independentment del seu desplaçament.

Els valors a afegir respecte a les coordenades de la finestra envoltant, que podem anomenar marges de seguretat, acostumen a ser constants del sistema de seguiment i imposen una restricció en la velocitat de l'objecte (canvi de posició acceptat durant el període de mostreig del sistema). Aquests marges de seguretat poden ser calculats de forma dinàmica, i depenen de:

- la velocitat aparent de l'objectiu, entesa com la variació de posició de la imatge sobre la retina, durant el període de mostreig del sistema de seguiment. Depèn per tant de la freqüència de mostreig del sistema i de l'òptica utilitzada.

- el model dinàmic de l'objectiu, que permeti fer una predicció més acurada del seu desplaçament entre imatges.

- la velocitat d'orientació de l'objectiu, ja que poden ser definits diferents els marges vertical i horitzontal.

- la possible variació de forma projectada de l'objectiu (o del propi objectiu), i per tant de la seva finestra envoltant.

És de gran interès que els marges de seguretat siguin tan petits com sigui possible, ja que d'aquesta forma les finestres de seguiment s'ajusten al seu objectiu i el solapament entre finestres és menys probable. A més a més disminuint l'àrea de les finestres s'aconsegueix rebaixar el temps de segmentació i càlcul del centroide, augmentant per tant la freqüència de mostreig. Hi ha aquí un punt important, i és que quan més ràpida sigui la freqüència de mostreig, més petita pot ser la finestra de seguiment i a la inversa.

Amb la intenció d'ajustar el màxim la mida de la finestra a la de l'objectiu alguns autors realitzen una estimació del tipus de moviment que presenta l'objectiu (normalment a través d'un filtre de Kalman) i apliquen una predicció de la nova posició de la finestra de seguiment que llavors pot ajustar-se millor a l'objectiu sempre que aquest no canviï de moviment [Frau,91]. El problema amb aquest mètode radica en la necessitat d'augmentar els marges de seguretat davant d'objectius maniobrables. Aquests marges poden donar lloc a finestres fins i tot més grans que les d'un sistema sense predicció si el moviment de l'objecte no és fàcil de modelitzar (pensem en una pilota que rebota de forma sobtada).

En el cas de aparèixer una superposició entre finestres el mètode proposat fracassa quan un objecte entra dins d'una altra finestra de seguiment, ja que s'agafen pel càlcul del centroide píxels que no pertanyen al mateix objecte. Apareix llavors una distorsió en la posició de les finestres que les arrossega a ajuntar-se (procés de fusió o fagocitació). Per solucionar aquest problema es poden utilitzar una extrapolació de trajectòria, tècniques de reconeixement dels objectius superposats, que permetin distingir-los, o bé imposar una limitació en la mida de la finestra de seguiment i associar els píxels segmentats a una única finestra de seguiment [Aranda,94].

2.4.3. Seguiment basat en el reconeixement

Els mètodes de seguiment basats en el reconeixement intenten superar els problemes presentats pel seguiment basat únicament en la segmentació, i donen més fiabilitat al procés de seguiment mitjançant una associació de dades relatives a l'objecte al llarg de la seqüència d'imatges [Lee, 93].

El seguiment basat en el reconeixement és en realitat una aplicació dels algorismes clàssics de reconeixement d'objectes. L'objecte és reconegut en imatges successives i la seva posició és estimada. D'aquesta manera s'aconsegueix la seva trajectòria.

L'avantatge d'aquest mètode és que poden ser calculades tant la translació com la rotació de l'objecte i com que l'objecte és conegut, és possible estimar la seva posició tridimensional a partir de la seqüència de projeccions i per tant la seva trajectòria 3D.

L'inconvenient obvi és que a cada nova imatge s'ha de realitzar el reconeixement de l'objecte seguit. Aquesta operació és considerada dins de la visió per computador d'alt nivell, degut a la seva complexitat intrínseca i a l'alt cost computacional associat a la seva resolució en la majoria de casos.

Per tant, l'eficàcia d'aquests sistemes de seguiment està limitada per l'eficiència del mètode de reconeixement i la seva velocitat de procés, que determinarà la latència del sistema de seguiment. Com que aquests dos paràmetres acostumen a ser contraposats, tots els sistemes de seguiment basats en el reconeixement agafen una solució de compromís entre la fiabilitat de l'algorisme de reconeixement i la seva velocitat, i limiten normalment el tipus d'objectes que poden ser reconeguts i les circumstàncies en que poden ser seguits. Totes les solucions presentades a la literatura disponible, intenten millorar la velocitat de procés a base d'utilitzar potents equips d'un cost econòmic elevat. El cost econòmic és un factor important a l'hora d'avaluar una solució ja que determina la seva possibilitat d'aplicació generalitzada. Aquests tres factors, fiabilitat en el reconeixement, velocitat i cost econòmic, marquen les diferències entre un sistema de seguiment i els altres.

Els sistemes de seguiment basats en el reconeixement trobats a la literatura poden ser classificats de la següent forma:

1. Mètode de correlació directa entre regions d'imatges.
2. Mètodes basats en característiques globals (com el test d'àrea i perímetre).
3. Mètodes basats en operadors de baix nivell.
    3.1. Línies rectes.
    3.2. Vèrtexs.

El mètode més directe de reconèixer un objecte en una altra imatge, consisteix en realitzar la correlació entre la zona d'imatge que conté l'objecte i tota l'altra imatge, o una zona en la que l'objecte és esperat (finestra de seguiment). El punt on la correlació és màxima és pres com la nova posició de l'objecte i així contínuament dins de la seqüència. L'avantatge d'aquest mètode és la seva simplicitat i per tant la seva possibilitat d'implemetar-ho mitjançant *hardware* específic [Brunelli, 95]. Entre els inconvenients destaquen:

- el volum de càlcul de la solució *software*.
- el mètode només permet translació de l'objecte. Els girs i canvis d'escala (zoom) no són tinguts en compte pel procés de correlació.
- els canvis que de forma sobtada pot tenir el fons de l'objecte quan es realitza el seu seguiment poden alterar el resultat de la correlació i fer perdre l'objecte.

Dintre dels treballs realitzats seguint aquesta tècnica destaca el processador d'imatge desenvolupat pel professor H. Inoue [Inoue,93], basat en una arquitectura paral·lela de correladors *hardware*, i que permet els girs i canvis d'escala dels objectes a base d'anar processant en paral·lel totes les correlacions amb la imatge patró, la imatge patró girada a esquerra i dreta, i aquestes augmentades i disminuïdes per un factor d'escala. En total són nou les correlacions que s'han d'efectuar en paral·lel. No evita el problema amb el fons, però tot i així els resultats obtinguts són molt positius i l'únic inconvenient que presenta el sistema és el seu elevat cost econòmic.

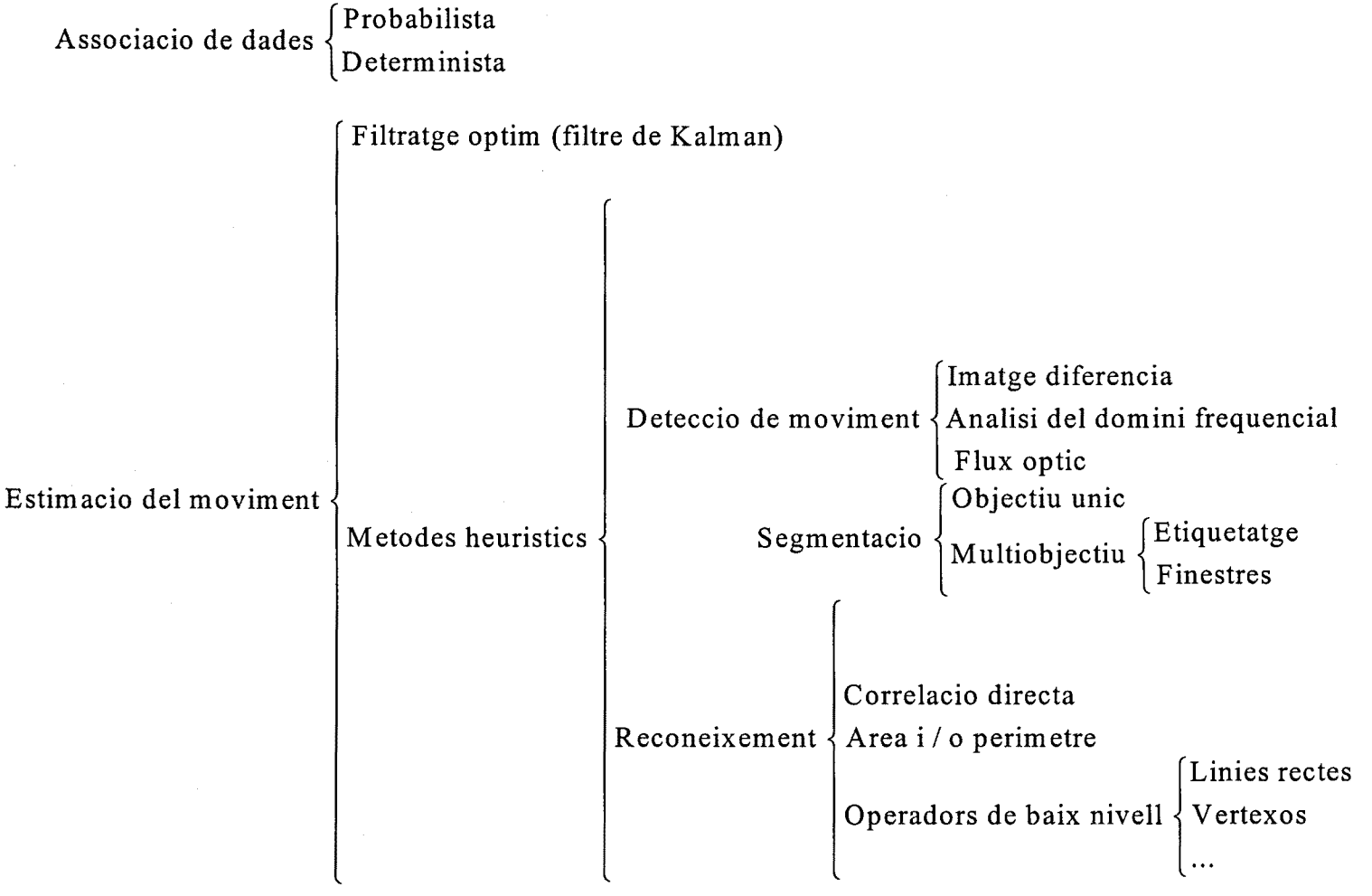
Els mètodes de test d'àrea i perímetre apareixen com una millora en la fiabilitat dels mètodes de seguiment basat en la segmentació. Recordem que en aquets mètodes l'estimació de la nova posició de l'objecte es feta a partir del càlcul del centroide. El nou centroide és associat a l'anterior només per proximitat. El càlcul d'àrea i el càlcul del perímetre són operacions fàcils que es poden realitzar de forma ràpida i que

permeten identificar els objectes. Units amb la restricció de la proximitat poden ser utilitzats com a filtre per donar més fiabilitat al seguiment, o per decidir en casos de seguiment de múltiples objectes. El principal inconvenient d'aquests mètodes és que l'objecte ha de ser seguit en la seva totalitat i lliure de solapaments amb altres objectes. Per altra banda la segmentació de l'objecte respecte al fons ha d'estar garantida durant tot el procés de seguiment.

El seguiment basat en operadors de baix nivell utilitza aquests per fer l'extracció d'algunes singularitats (*features*) de la imatge, com ara vèrtexs [Wang, 92] [Tan, 93] [Smith, 95] [Wang, 95] o línies rectes [Deriche, 90] [Jeng, 91] [Mirmehdi, 93] [Lai,93], que poden ser fàcilment identificables dins d'una seqüència [Aggardwal, 81] [Feddema, 90] [Krishnan, 95]. Cadascuna d'aquestes singularitats va acompanyada d'un conjunt d'atributs (*token*) que permeten la seva distinció respecte dels altres punts característics. El següent pas és la solució de l'anomenat problema de la correspondència entre els punts característics que pertanyen a imatges consecutives. Aquest problema és considerat costós ja que calen N^2 comparacions sent N el nombre de punts característics. Per últim, una vegada identificades les correspondències, s'estima el moviment de l'objecte a partir de les successives posicions dels punts singulars dins de la seqüència d'imatges. A partir d'aquestes dades de moviment també és possible conèixer l'estructura de l'objecte (*structure from motion*) [Tomasi, 91] [Jerian, 91] [Lee, 93] [Pei,94].

Amb l'objectiu de minimitzar el temps de càlcul es sol recórrer a preprocessadors especialitzats per detectar les singularitats i els seus atributs. El procés d'associació de dades es pot restringir a un àrea d'interès dins de la imatge anomenada finestra de seguiment i que actua com a restricció en la velocitat dels punts seguits. També es poden utilitzar arquitectures paral·leles ja que el procés d'associació pot ser efectuat de forma múltiple amb diferents singularitats. L'aspecte més important resideix sense dubte en l'estudi de quins són els atributs més interessants a l'hora de garantir un reconeixement fiable i en quin format són codificats de manera que l'associació pugui ser més ràpida.

2.5. Esquema dels diferents mètodes d'estimació de moviment mitjançant visió per computador



CAPÍTOL 3

3. Sistema de seguiment proposat

3.1. Introducció

En el present capítol es descriu el sistema proposat per localitzar i seguir objectes dins d'una escena mitjançant visió per computador. El sistema proposat pertany al tipus de seguiment basat en el reconeixement, tipus descrit al capítol anterior.

D'entre les diferents metodologies que es segueixen dins del seguiment basat en el reconeixement, s'ha elegit la de considerar únicament aquelles regions de la imatge que presenten alguna singularitat (característiques locals). D'aquesta forma es redueix la quantitat d'informació a processar durant l'aparellament entre imatges, tant en el cas de l'estereovisió com en el cas del seguiment dins d'una seqüència.

A continuació es detallen els objectius que s'han abordat.

3.1.1. Objectius del sistema proposat:

- La detecció i localització de les característiques locals dels objectes que apareguin en l'escena.
- La valoració i selecció d'aquestes característiques locals.
- La identificació i reconeixement fiable de les característiques locals davant de possibles translacions i girs dins de la seqüència d'imatges.
- L'aparellament de les característiques locals detectades, basat en un reconeixement local de la imatge. S'han d'abordar dos objectius:
    - estèreo: aparellament de característiques entre dues imatges de la mateixa escena amb diferent perspectiva.
    - seguiment: aparellament de característiques entre imatges consecutives.
- La validació dels aparellaments a partir de l'anàlisi de coherència de les trajectòries tridimensionals de les característiques locals seguides.
- L'optimització del sistema de reconeixement local utilitzat, amb el criteri cost-eficiència-temps per a una implementació *hardware* que permeti l'operació en temps real.

3.1.2. Descripció del sistema proposat.

El sistema proposat està compost de tres mòduls de processat de dades que operen en dues fases: la fase d'inicialització i la fase de seguiment. A continuació es dona una descripció de les funcions dels diferents mòduls de processat.

Mòdul de processat d'imatge:

Per fer la detecció d'aquestes característiques locals de la imatge (*local features*), s'ha recorregut al disseny i implementació d'un processador d'imatge específic. El processat realitzat consisteix en:

a)L'adquisició i emmagatzematge de les imatges procedents de les cameres esquerra i dreta.

b)L'extracció dels contorns de la imatge. (Pot ser substituït per un binaritzat quan les imatges són prou senzilles o en entorns industrials on es pot regular la il.luminació). Aquest preprocessat, de fet, pot estar basat en qualsevol forma de segmentació que doni una imatge binària de sortida, incloent color, textura o imatge diferència per exemple.

c)La descripció local de la imatge obtinguda per a cada pixel a fi de poder realitzar el seu posterior reconeixement. Aquesta descripció actua com a *token* o vector característic que identifica la regió que envolta al pixel i permet la seva distinció respecte dels altres durant el seguiment

Mòdul d'inicialització:

a)Detecció i selecció automàtica de les característiques locals presents en una de les imatges, bé en la totalitat de la imatge o bé en finestres definides per l'usuari.

b)Aparellament dels punts d'interès seleccionats amb la imatge procedent de l'altra camera. Nova selecció basada en la fiabilitat de l'aparellament i la distància.

En acabar aquesta fase el sistema disposa de la informació de l'estat corresponent a cadascuna de les característiques locals (posició i vector característic (***descriptor***) dins de cada imatge).

Mòdul de seguiment:

a)Seguiment bidimensional: De forma independent per la imatge esquerra i la imatge dreta es realitza la cerca i aparellament de les característiques locals en la següent imatge al voltant de la posició que ocupaven en la imatge anterior (o la posició donada per la fase d'inicialització). També s'obté una mesura de fiabilitat de l'aparellament entre vectors característics.

b)Seguiment tridimensional: A partir de la informació procedent del seguiment bidimensional:

b.1) es calcula la trajectòria tridimensional de cada objectiu.

b.2) s'efectua una anàlisi i supervisió de les dades amb la finalitat de filtrar possibles errors d'aparellament. S'analitza la coherència de la trajectòria.

b.3) es realitza una estimació de l'estat de les característiques locals en la següent mostra (posició, velocitat i vector característic dins de cada imatge). Apareixen llavors dues opcions:

- Realimentar el mòdul d'aparellament amb el nou estat de les característiques locals.
- Si el nombre de punts seguits no és prou elevat, degut a una incoherència en les trajectòries, tornar a la fase d'inicialització.

A la figura 3.1. es presenta un diagrama general del sistema de seguiment proposat en aquesta tesi. A continuació es descriuen, a cada apartat d'aquest capítol, les solucions proposades per cadascuna de les parts que composa el sistema.

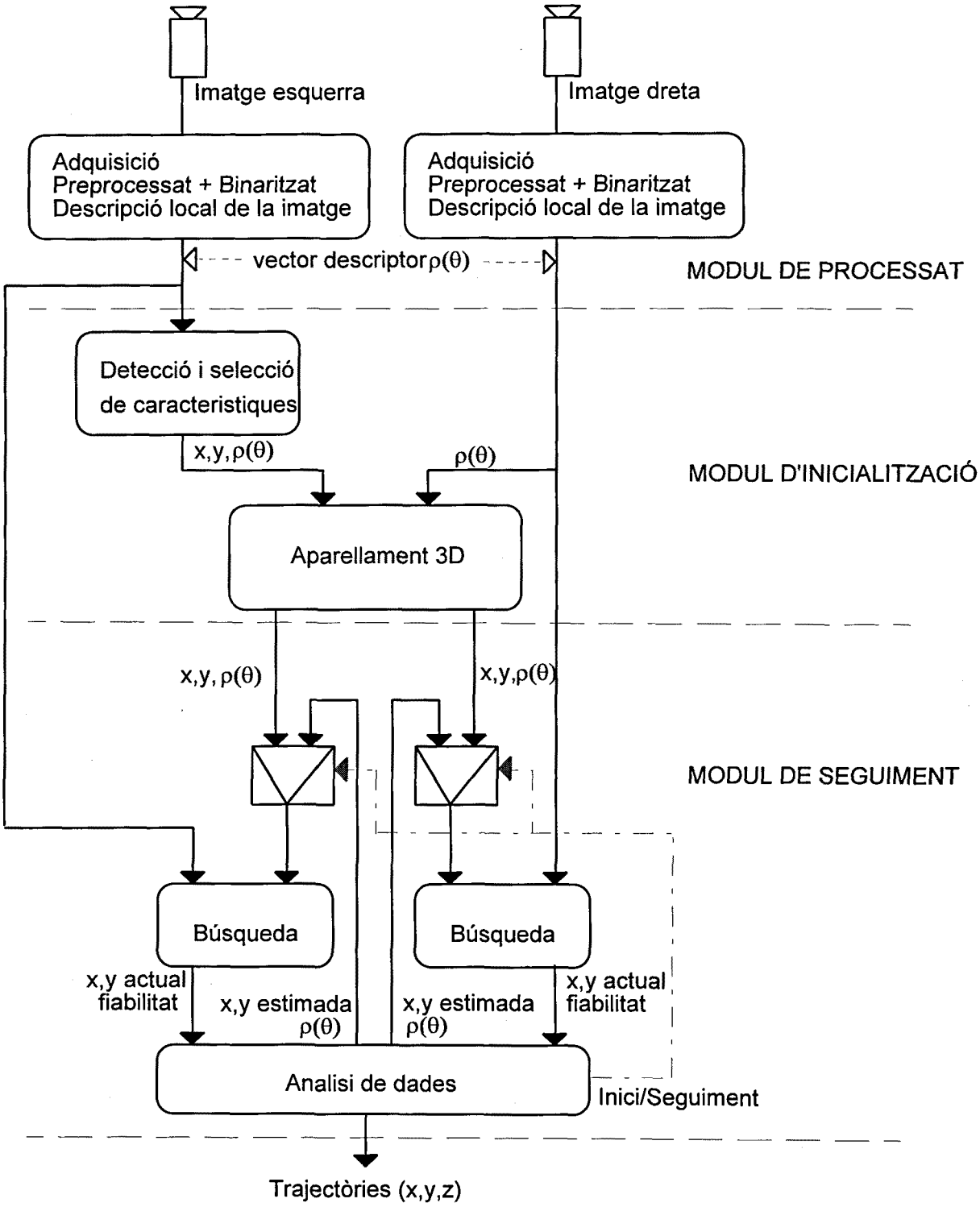


Figura 3. 1. Diagrama general del sistema de seguiment tridimensional proposat

3.2. Extracció de les característiques locals de la imatge

3.2.1. Definició i caracterització

Per poder fer el seguiment d'un element de la imatge, cal en primer lloc que aquest element pugui ser "reconegut" dins de la imatge. Això vol dir que presenti algunes característiques discriminants que el facin singular o diferent respecte al seu entorn. D'aquesta forma serà possible la seva segmentació, reconeixement i localització posterior.

D'entre les diferents metodologies per segmentar les imatges s'ha decidit aplicar l'extracció de contorns sobre la imatge monocolor. D'aquesta forma aconseguim una imatge binària que suposa una gran reducció de la quantitat d'informació a processar, sense perdre la informació que és rellevant pel seguiment: els límits dels objectes o formes que apareguin a l'escena [Rao,91]. Cal dir que no és l'única forma de preprocessat que permet el sistema. De fet, qualsevol forma de segmentació que ofereixi una imatge binaria de sortida pot ser aplicat (binarització, color, textures...).

Però no tots els pixels de contorn poden ser identificats respecte a la resta. Aquest seria el cas dels pixels del contorn d'una línia recta o amb poca curvatura, els quals són impossibles de seguir ja que tenen veïns amb la mateixa informació característica. Cal que el contorn presenti un canvi de direcció si el volem localitzar en les dues dimensions del pla de la imatge. Aquest seria el cas dels vèrtexs, cantonades i unions de segment per exemple. També els contorns tancats de petita mida (com els forats) són regions clarament identificables.

3.2.2. Descripció local de la imatge

En aquest apartat es presenta un mètode per detectar i identificar aquestes petites regions singulars de la imatge, que constitueixen el conjunt de característiques locals que seran seleccionades per al seu posterior aparellament i seguiment. Aquest mètode està basat en la signatura polar del contorn dins d'una determinada zona de la imatge.

Una signatura és una representació funcional unidimensional d'un contorn. Existeixen moltes formes de generar una signatura, però independentment de com es generen, la idea bàsica consisteix en reduir la representació del contorn de dues dimensions a una [Casals,83][Villà,83][Gonzalez,87]. D'aquesta forma és més fàcil descriure el contorn i es simplifica la seva identificació.

Una de les formes més simples de descriure el contorn d'una figura és donar la distància des del centroide del contorn fins a cadascun dels seus punts com a funció de l'angle, $\rho(\theta)$, tal com mostra la figura 3.2. El mètode és equivalent a realitzar la transformada polar dels pixels de contorn respecte al centroide.

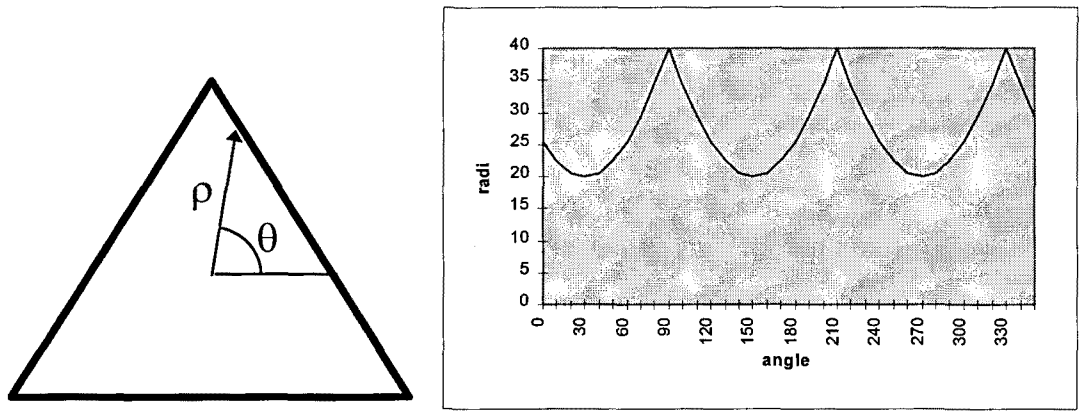


Figura 3. 2. Transformada polar d'un contorn

Aquest mètode permet una normalització de la mida, de la posició i de l'orientació del contorn de la figura, per la qual cosa ha estat utilitzat per molts autors com a pas previ al reconeixement d'objectes [Amat,89][Jeng,91][Sekita,92][Friedland,92].

En el cas que la figura tingui concavitats pot passar que per un determinat valor de angle, θ, hi hagi més d'un pixel de contorn a diferent distància (radi) (veure fig. 3.3). En aquest cas s'acostuma a utilitzar pel valor de la signatura el valor del radi més petit, o sigui, la transformació del pixel de contorn més proper al centre de la figura.

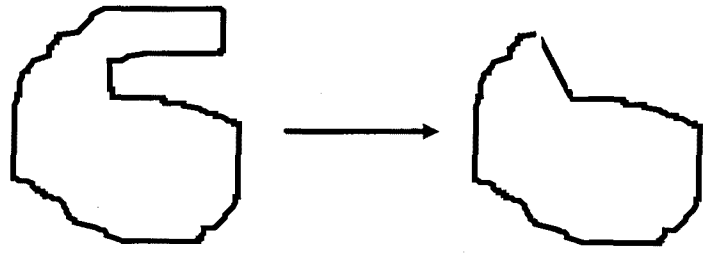


Figura 3. 3. Distorsió produida per la transformació polar en el cas de contorns amb concavitats.

La transformació pot ser definida igualment per contorns oberts, o que cauen fora de la imatge o l'espai de transformació, assignant un valor màxim als radis per totes aquelles direccions en les que no es troba cap pixel del contorn.

El que es pretén però, no és estudiar la transformació polar de la totalitat del contorn de la figura, si no únicament la transformació d'aquelles regions del contorn que, per la seva singularitat en les dues dimensions del pla de la imatge, siguin representatives de la figura i permetin el seu seguiment de forma fiable. D'aquesta

manera farem una reducció dràstica de la quantitat de característiques locals a aparellar i seguir, amb la qual cosa obtindrem un temps de procés molt més baix.

La transformació polar serà aplicada d'una forma local a cada punt de la imatge. Aquesta signatura no sols pot servir per identificar una regió de la imatge i així permetre el seu reconeixement, si no que a més a més dona informació en un format adient per trobar característiques locals de la imatge amb els contorns.

3.2.3. Discretització i optimització de la transformació polar.

Donat el caràcter discret de la imatge, la codificació polar de la regió singular serà també discreta amb una limitació en la mida dels radis i en el nombre de radis o quantitat de mostres del contorn (angles). Apareixen doncs dues qüestions bàsiques:

- quina ha de ser la mida mínima (en pixels) de la regió estudiada.
- quantes mostres del contorn són necessàries per identificar-lo.

Aquestes dues preguntes s'han de resoldre tenint en compte que cal buscar un compromís entre la màxima fiabilitat durant el reconeixement i el mínim cost computacional.

S'ha considerat que les imatges seran mostrejades amb pixels quadrats i s'ha pres el criteri de considerar les regions d'estudi (finestres) de forma quadrada, de forma que les dues coordenades de la regió tenen la mateixa longitud, N. La longitud dels costats de la regió analitzada ha de ser senar ja que d'aquesta forma els radis de la transformació seran simètrics respecte al pixel central de la regió. Aquest pixel central no serà considerat al efectuar la transformació. És a dir, si els radis tenen una longitud màxima de ρ_{max} pixels, la mida del costat de la regió analitzada queda determinat per l'expressió: $N=2\rho_{max}+1$.

Així doncs només ens cal definir quina serà la longitud màxima dels radis. Aquesta distància des del centre de la finestra fins al pixel de contorn (ρ), haurà de ser codificada utilitzant el menor nombre de bits possible, en vista a una implementació *hardware* de la transformació polar. Cal assignar un valor de la codificació (diguem el valor més alt) corresponent a la possibilitat de que en una determinada direcció no es trobi contorn. Per tant la longitud màxima dels radis, ρ_{max}, serà 2^n-1, essent n el nombre de bits de codificació dels radis.

Ara el nombre de bits de codificació dels radis (n) determina les mides possibles de la regió a analitzar, de tal manera que $N=2\rho_{max}+1$ pot ser substituïda per l'expressió $N= 2^{n+1}-1$, pels diferents valors que pugui tenir n. Finestres de dimensions intermitjes, entre un valor n i $n+1$, tindrien un cost de codificació de radis equivalent a les finestres de dimensió superior (n +1 bits) i donarien menys informació.

De la mateixa manera el nombre *n* determina el nombre màxim de radis (mostres) que podrem considerar dins de la regió. El nombre de radis màxim en una regió de NxN pixels coincideix amb el nombre de pixels de la perifèria d'aquesta regió i ve donat per l'expressió 4(N-1). Substituint N a la fórmula ens queda que el nombre màxim de radis és igual a $8(2^n-1)$.

Tots aquest resultats queden resumits a la següent taula:

Nº de bits (n)	Radi màxim ($\rho max=2^n-1$)	Finestra (NxN) ($N=2^{n+1}-1$)	Nº de pixels de la finestra	Nº max de radis= $8(2^n-1)$
1	1	3x3	9	8
2	3	7x7	49	24
3	7	15x15	225	56
4	15	31x31	961	120
5	31	63x63	3969	248

Taula 3.1. Relació entre nombre de bits de codificació, dimensió de la finestra de transformació i nombre màxim de radis.

Degut al caràcter local que tenen les singularitats, i de cara a reduir el temps de càlcul, convé que l'àrea analitzada sigui tant petita com es pugui. El cost de processat és proporcional a l'àrea analitzada, això és N^2 (columna 4 de la taula).

Les finestres de 3x3 queden descartades ja que no donen prou informació per estimar si un contorn és o no localment recte [Oliver,93].

En el cas de les finestres de 7x7, tot i que la quantitat de mostres radials pot ser elevada, fins a 24, el marge de valors que poden agafar queda molt reduït, tant sols als valors 1, 2 ó 3. Això fa que la transformació polar sigui molt sensible al soroll present en la localització dels pixels de contorn. La variació d'un sol pixel provoca un 33% d'error de mesura en el radi corresponent en el millor dels casos (valor del radi igual a 3).

Aquesta excessiva sensibilitat al soroll provoca uns canvis en la codificació que dificulten enormement l'associació de dades necessària pel posterior procés de seguiment. La "qualitat" de la transformació pot ser mesurada com la relació entre el rang de valors i un possible error d'un pixel en la mesura (error de resolució).

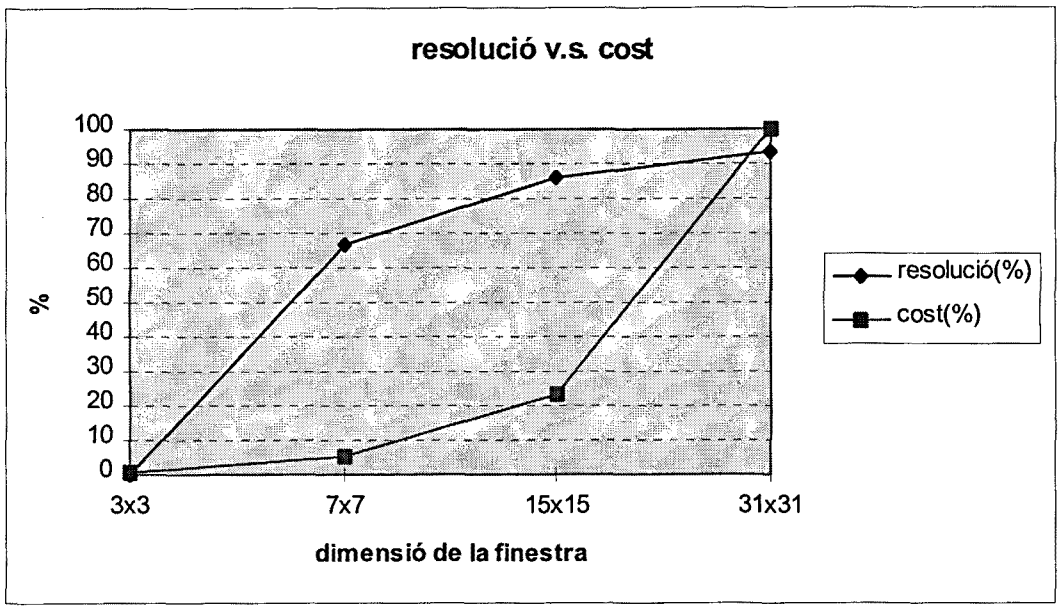


Figura 3. 4. Augment de la resolució i el cost en funció de la dimensió de la finestra de transformació polar.

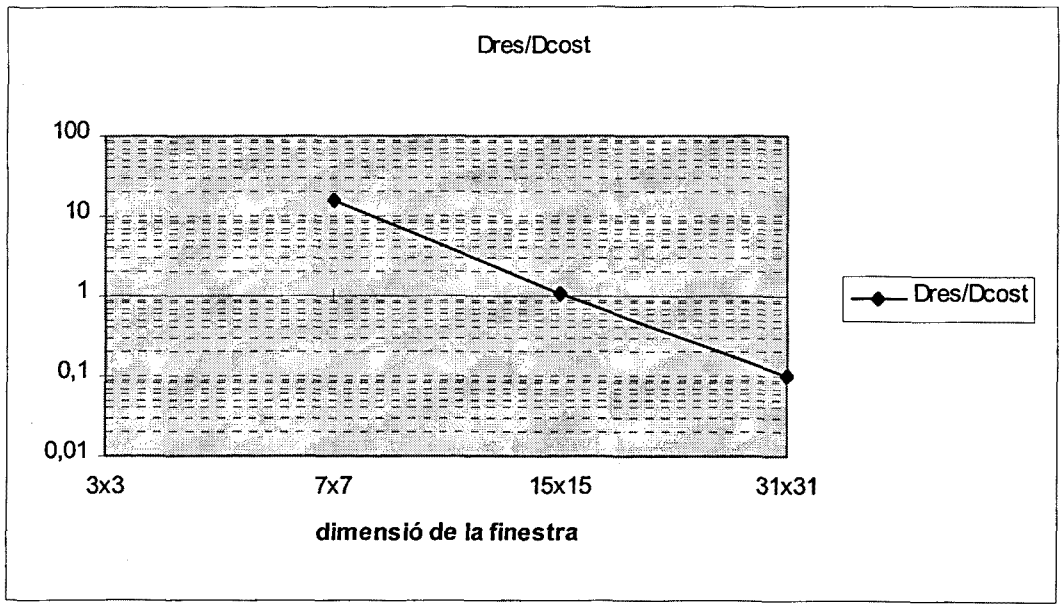


Figura 3. 5. Relació increment de resolució (Dres) v.s. increment de cost (Dcost) en funció de la dimensió de la finestra de transformació polar. Els resultats es mostren en escala logarítmica.

El que s'ha fet és buscar la dimensió de finestra òptima que millori la qualitat de la transformació polar en front del soroll, però que no suposi un augment del cost de computació desproporcionat. L'estudi s'ha efectuat sobre la relació: increment de resolució / increment de cost. (veure figures 3.4 i 3.5).

La solució adoptada ha estat considerar finestres de 15x15 pixels. El ventall de valors possibles pels radis s'amplia així de 0 a 7 i l'error produït per la variació en un pixel degut al soroll, queda reduït a un 14% del valor de la mesura.

La transformació polar en finestres de mida superior resulten d'un cost massa elevat tant en temps per a la solució *software*, com en components per a la solució *hardware* (necessitat de disseny en VLSI). De fet l'augment d'un sol bit en la codificació dels radis (duplicant el rang de valors) quadruplica el cost de la solució ja que la transformació és bidimensional. L'augment aconseguit en la resolució no justifica el cost adicional. A més a més, amb finestres de 31x31 pixels i superiors es perd el caràcter local de la mesura.

Val la pena comparar aquest resultats amb els obtinguts per [Moravec,77] que utilitza finestres de 5x5 fins a 11x11 per localitzar punt singulars dins de la imatge. També [Oliver,93] limita a 7x7 els patrons utilitzats per la localització de vèrtexs.

Tal com mostra la taula 3.1, el nombre màxim de radis que es poden considerar en una finestra de 15x15 és igual a 56. Ara bé, suposant que el contorn a dins de la finestra de transformació és continu, un nombre de mostres més petit donaria prou informació. El nombre de radis ha ser múltiple de 8 si volem que les mostres estiguin distribuïdes de forma homogènia al voltant del centre (degut a les simetries vertical, horitzontal i diagonals que té la finestra). Una reducció en el nombre de radis porta aparellada una reducció en la quantitat d'informació associada a cada punt de la imatge. Això representa una reducció en la memòria i el temps (o cost de la implementació *hardware*) necessari per processar després aquesta informació.

S'ha de tenir en compte que els radis que arriben fins als pixels perifèrics de la finestra de transformació tenen un solapament molt elevat en els pixels centrals de la finestra, és a dir, els pixels centrals donen informació redundant respecte la seva distància al centre. Aquesta redundància en la informació és més elevada quan més ens acostem al centre de la finestra de transformació (figura 3.6).

1	1	1	1	1	1	1
1	1	2	1	2	1	1
1	2	3	3	3	2	1
1	1	3	0	3	1	1
1	2	3	3	3	2	1
1	1	2	1	2	1	1
1	1	1	1	1	1	1

1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	2	1	1	1	1	1	2	1	1	1	1
1	1	1	1	2	2	1	1	1	2	2	1	1	1	1
1	1	1	1	2	2	2	1	2	2	2	1	1	1	1
1	2	2	2	3	2	2	3	2	2	3	2	2	2	1
1	1	2	2	2	3	4	3	4	3	2	2	2	1	1
1	1	1	2	2	4	7	7	7	4	2	2	1	1	1
1	1	1	1	3	3	7	0	7	3	3	1	1	1	1
1	1	1	2	2	4	7	7	7	4	2	2	1	1	1
1	1	2	2	2	3	4	3	4	3	2	2	2	1	1
1	2	2	2	3	2	2	3	2	2	3	2	2	2	1
1	1	1	1	2	2	2	1	2	2	2	1	1	1	1
1	1	1	1	2	2	1	1	1	2	2	1	1	1	1
1	1	1	1	2	1	1	1	1	1	2	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Figura 3. 6. Nombre de solapaments que es produeixen a cada pixel considerant els 24 radis d'una finestra de 7x7 pixels i els 56 radis d'una finestra de 15x15 pixels.

Degut al solapament, l'augment del nombre de radis (i per tant del cost) no provoca un augment lineal de la quantitat d'informació processada (quantitat de pixels mostrejats dins de l'àrea) tal i com es pot veure a la figura 3.7.

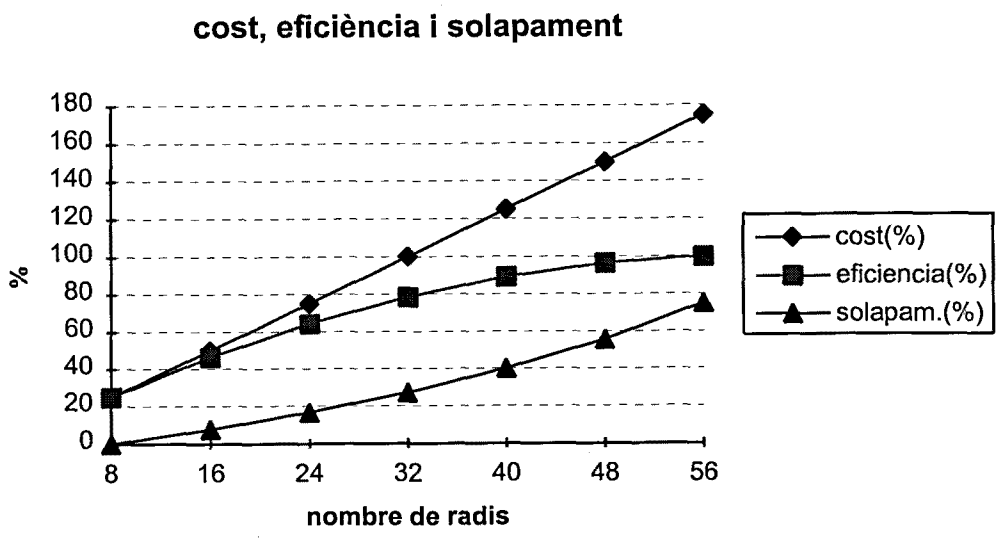


Figura 3. 7. Cost, eficiència i solapament de l'operador en funció del nombre de radis.

La determinació del nombre mínim de radis ha estat basada en proves empíriques que demostraven l'eficàcia o no de la solució. Les proves realitzades amb quatre radis per exemple, van demostrar la seva escassa fiabilitat davant de rotacions de l'objecte dins de l'àrea d'imatge transformada.

S'ha realitzat un estudi comparatiu amb transformacions de 8 i de 16 radis a fi de determinar la fiabilitat de la identificació. El patró utilitzat en el cas de 8 radis ha sigut l'indicat a la figura 3.8:

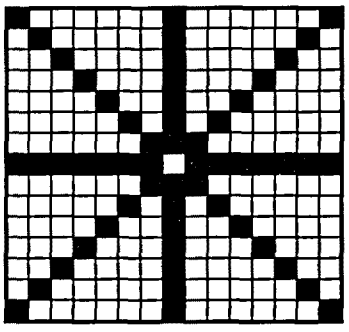


Figura 3. 8. Patró utilitzat per la solució de 8 radis

Per ampliar aquest patró a 16 radis ens trobem amb dues solucions possibles en funció del pixel del contorn que vulguem mostrejar [Bresenham,65]. Qualsevol de les dues solucions però, no compleix la condició d'homogeneïtat en la distribució dels radis, la qual cosa provoca certes imprecisions a l'hora d'estimar un possible gir del contorn transformat.

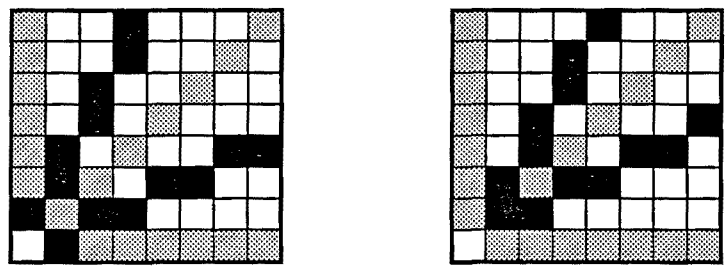


Figura 3. 9. Dues possibilitats per ampliar el patró a 16 radis. (Només es mostra un quadrant)

El resultat de les proves realitzades amb diferents característiques locals (amb contorns oberts i tancats) ha demostrat que la millora apreciada amb la solució de 16 radis no és significativa comparada amb la reducció de cost aconseguida amb la solució de 8 radis (la meitat).

Els resultats de mostrejar diferents figures amb 8 i 16 radis poden ser comparats a l'annex 1. Com és pot apreciar en la majoria dels casos (per figures relativament senzilles) el valor dels 16 radis es pot treure per interpolació del valor del altres 8 radis. A l'annex, la longitud dels radis diagonals no ha sigut compensada (multiplicant per $\sqrt{2}$) en la transformació de la columna dreta per facilitar la seva correspondència amb el gràfic de la columna esquerra.

A l'annex 2 es pot apreciar com la solució de 8 radis es capaç de discriminar prou bé les diferents figures amb el mètode de reconeixement que serà descrit a l'apartat 3.4 d'aquest capítol. Aquests resultats demostren que 8 radis són suficients sempre que els contorns no continguin variacions de direcció massa brusques dins de la finestra de convolució (o sigui que la seva transformació sigui prou homogènia).

3.3. Selecció de les regions singulars.

En aquest apartat es presenta un mètode per obtenir un valor representatiu de la fiabilitat amb la que una característica local de la imatge podrà ser identificada en el posterior procés de seguiment. Aquest mètode es basa en la informació continguda en la transformació polar descrita anteriorment.

Per a la detecció de característiques locals dins d'una imatge és ampliament utilitzada la convolució d'una màscara patró sobre cada punt de la imatge. La màscara es considerada més o menys bona en funció de la seva capacitat de discriminar la característica buscada de la resta de la imatge. Després la comparació del valor resultant de la convolució amb un llindar fixat determina la singularitat o no del punt en qüestió.

El que es pretén no és proporcionar un operador binari que ens indiqui si una determinada regió de la imatge és o no singular, sinó una funció que doni un *índex de singularitat.* Aquest índex, qualitatiu, ens ha de permetre seleccionar aquelles regions de la imatge que presenten una singularitat relativa més elevada, entenent que d'aquesta forma el seu seguiment serà més fiable. Amb aquest index la selecció pot ser ajustada de forma automàtica a les característiques de cada nova escena. D'aquesta forma es redueix el problema que tenen els detectors binaris amb la selecció del llindar i que provoca que segons les circumstancies siguin massa o poc sensibles a les caracteristiques locals presents en l'escena.

Habitualment els operadors que fan referència a l'interes d'un determinat punt de la imatge operen sobre imatges multinivell tot aplicant operadors de cost computacional elevat i/o que donen resultats després de varies iteracions [Moravec,77] [Rosenthaler,92].

En el nostre cas hem obtingut la singularitat d'una determinada regió de la imatge a partir de la mateixa informació que servirà per realitzar el seguiment d'aquesta característica local en quadres succesius. S'ha deduït un mètode que és prou fiable i també ràpid ja que sols necessita una iteració per la imatge.

Amb la intenció de treure el màxim profit de la informació de la que es disposa, s'ha estudiat un mètode heurístic que valori la singularitat d'un contorn tant en el cas que es tracti d'un contorn tancat dins de l'àrea transformada (15x15 pixels), com si en cas contrari, el contorn surt d'aquesta àrea i per tant és vist com un contorn obert.

A continuació es fa una enumeració dels diferents mètodes estudiats i valorats a fi de determinar aquelles característiques locals de la imatge de contorns que mostren un *index de singularitat* més elevat.

3.3.1. Anàlisi de la segona derivada

Un primer mètode avaluat per determinar la singularitat dels pixels d'un contorn ha sigut l'estudi analític de la transformada polar i les seves derivades. A partir de la transformada polar d'un contorn poden ser detectats els punts singulars per un canvi brusc en el creixement de la funció, o dit d'una altra forma per una discontinuïtat en la seva primera derivada.

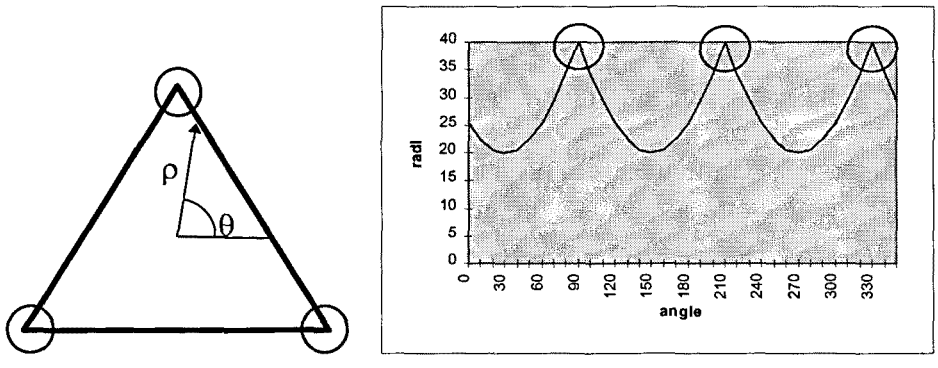


Fig.3.10. Detecció dels punts singulars d'un contorn mitjantçant la seva signatura.

La discontinuïtat pot ser trobada a través de la comparació del valor absolut de la segona derivada amb un llindar determinat ($|\rho''(\theta)| > \text{llindar}$). Aquest valor per si mateix podria servir com una estimació de la singularitat d'aquest pixel.

Aquest mètode funciona satisfactòriament quan la funció transformació $\rho(\theta)$ és continua o té una alta resolució en les seves dues dimensions. El problema es presenta amb una discretització de baixa resolució com succeeix amb la nostra representació polar de 8 radis i amb 8 valors possibles. La baixa resolució d'aquesta funció fa que aquest mètode sigui poc fiable, fins i tot intentant fer una interpolació amb les poques dades disponibles, tal i com s'ha pogut comprovar experimentalment amb diferents tipus d'imatges.

D'altra banda existeixen a la imatge altres característiques locals no tingudes en compte pel mètode de la segona derivada. Tal és el cas dels forats petits (de multiples formes) que cauen dins de la finestra de transformació. En aquests casos la transformació es realitza sobre un contorn tancat.

3.3.2. Anàlisi de la homogeneïtat en el valor dels radis

Amb aquest criteri es preten donar més pes a aquelles regions de la imatge que presenten una semblança en els valors dels radis de la seva transformació polar (com és el cas de les circumferències), i per tant es mostren invariants a la rotació de la imatge. Degut a la discretització del mostreig radial del contorn, les transformades polars amb valors dels radis molt diferents són poc estables davant de rotacions petites de la característica seguida.

Un paràmetre que ens permet avaluar la dispersió en les mostres radials es la variança mostral. Es tractaria dons de minimitzar aquest paràmetre. El resultat del càlcul de la variança però, mostra grans diferències en funció de la grandària dels valors de les mostres, no actuant igual davant de la mateixa dispersió absoluta quan les mostres tenen un valor alt que quan aquestes tenen un valor baix (circumferències de diferent radi). Per compensar aquesta dificultat s'ha utilitzat un valor normalitzat de la dispersió dels valors dels radis mitjançant el paràmetre *VN=var(X)/mitja(X)*, sent *X* el conjunt de mostres que formen els valors dels vuit radis de la nostra transformació polar.

Això permet la identificació de circumferències de petit radi (entre 1 i 7 pixels) així com d'altres contorns tancats de formes ben diferents. En aquests casos el valor del paràmetre VN és representatiu de la dificultat de fer el seguiment del contorn davant d'un gir de l'objecte. Per una circumferència, de qualsevol radi, continguda a dins de la finestra de transformació obtindriem un valor de VN igual a zero.

Amb els contorns oberts, o sigui, amb aquells objectes que per la seva grandària només una petita part del seu contorn pot ser transformat, només s'ha d'estudiar la dispersió dels valors dels radis que arriben fins a un pixel contorn (que tenen un valor entre 0 i 6). Els radis que no toquen cap pixel de contorn agafen el valor màxim de la codificació (7) i no dònen cap informació relevant.

Els contorns oberts presenten en general una fiabilitat més baixa que els contorns tancats degut a que tenen alguns radis amb valor igual a set. Per estudiar els resultats d'aquestes proves, els contorns oberts han sigut modelitzats com un vèrtex d'angle variable (α, entre 10 i 180 graus), que es trobi a diferents distàncies respecte al centre de la finestra de trasformació (d, entre 1 i 7 pixels). (Figura 3.11).

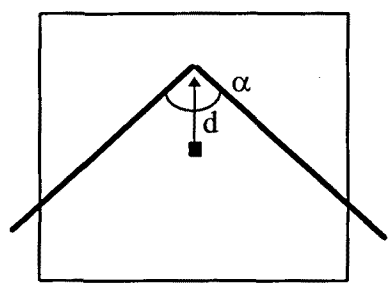


Figura 3.11. Modelització d'un contorn obert com un vèrtex d'angle (α) variable a distància (d) també variable respecte el centre de la finestra de transformació polar.

La dispersió radial (VN) del diferents vèrtexs modelitzats ha sigut mesurada i es presenten els resultats a les gràfiques de la figura 3.12.

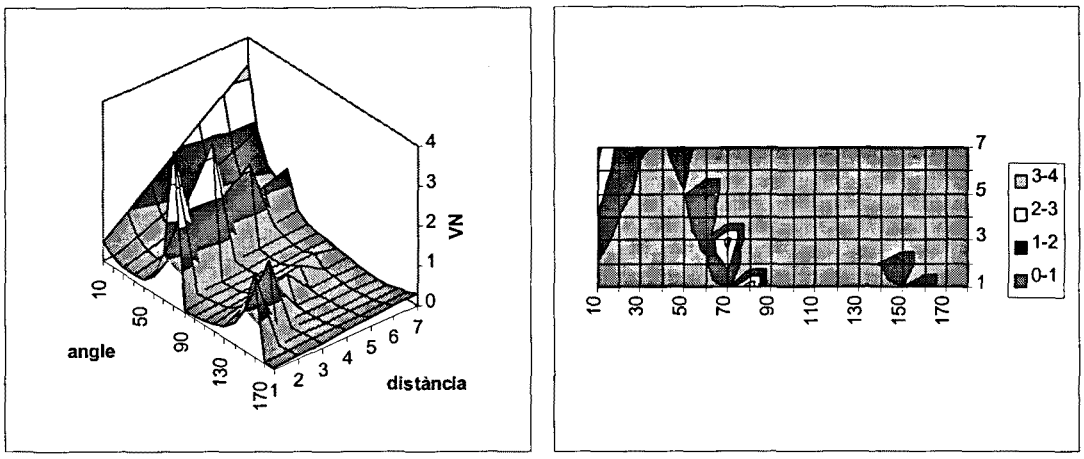


Figura 3.12. Dispersió dels valors radials dels contorns oberts. VN=var(X)/mitja(X)

Com es pot veure en aquestes gràfiques, la homogeneïtat dels radis és un paràmetre que pot servir per fer l'estudi de la fiabilitat de caracterìstiques locals amb contorns tancats, pero que no discrimina diferents contorns oberts. Apareixen uns pics per alguns valors concrets de parelles (α,d) com és el cas de la parella (α=70° , d=3pixels) o la parella (α=80° , d=1pixel). Aquests pics són deguts a la limitada quantitat de radis de la transformació polar utilitzada (vuit). Això fa que els vèrtexs quan augmenten d'angle passen de forma brusca de tallar un radi a no tallar-lo. Es pot observar que els pics apareixen a diferents angles en funció de la distància al centre de la finestra, però guardant una relació lineal entre les dues variables.

S'han realitzat diferents proves amb aquest criteri de forma que s'ha pogut constatar la seva escasa utilitat per discriminar caracteristiques locals de l'escena.

Ha sigut provada, també sense èxit, una funció que maximitzés per una banda la homogeneitat dels radis i de l'altre la distància a un patrò model (circumferència de radi igual a 3,5). En aquest cas la funció a minimitzar és : ***D*VN.*** (veure la figura 3.13).

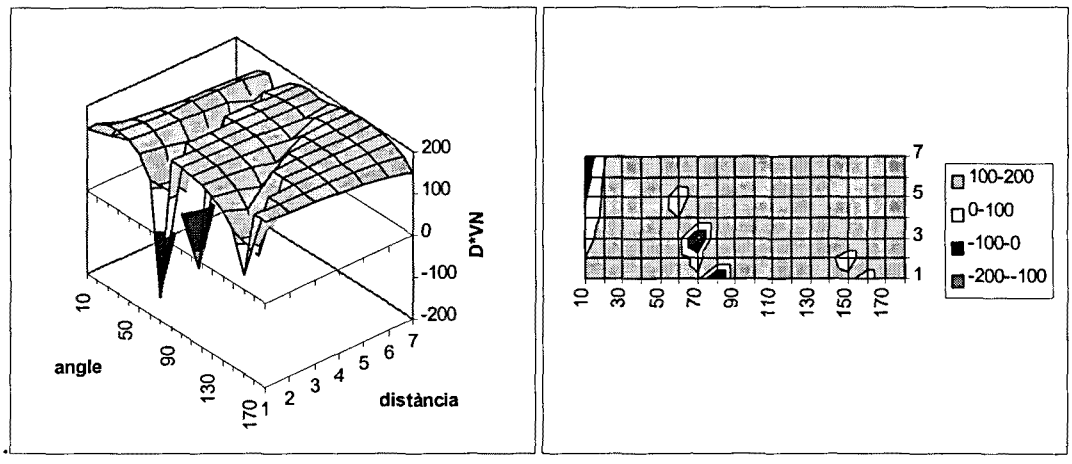


Figura 3.13. Anàlisi combinat de la dispersió radial i de la comparació amb un mòdel per contons oberts.

3.3.3. Comparació amb un patró model

A partir de l'anàlisi de la transformació de determinades regions de la imatge que a priori podem interpretar com característiques locals, s'han establert una sèrie de condicions que cal considerar a l'hora d'avaluar la fiabilitat de les regions en el procés de seguiment.

Aquestes condicions són:

- *Contorns tancats:* Tal com ha estat definida la codificació dels radis de la funció $\rho(\theta)$, el valor d'un radi quan no troba cap contorn en la seva direcció és forçat a set (valor màxim). Això implica de forma evident que un contorn tancat és més probable d'identificar ja que disposem de més radis amb informació.

- *Invariança a la rotació:* Si es tractés d'un contorn tancat amb el mateix valor en tots els radis de la transformació, com seria el cas d'una circumferència al voltant del centre de àrea de transformació, llavors la fiabilitat en el seguiment seria màxima donada la invariància a la rotació en el valor de les mostres radials.

- *Invariança a l'escala*: D'entre totes les circumferències, aquelles que tenen un radi proper al limit de la codificació (0 ó 7) presenten molta inestabilitat davant del soroll en la localització del contorn o davant d'un apropament (allunyament) de l'objecte durant el seu seguiment.

Amb aquests criteris s'ha agafat com a model òptim de caracteristica local una circumferència de radi igual a 3,5 pixels (valor mig dels radis) i s'ha definit una funció fiabilitat de seguiment d'una regió de la imatge, com la que expressa la proximitat de la seva transformació amb la transformada d'aquest model de referència. D'aquesta forma es defineix la fiabilitat F com:

$$F = MAX\_D - D \qquad \text{[Eq. 3.1]}$$

on D és una mesura de distància donada per alguna funció que maximitzi els anteriors criteris i MAX_D és el valor màxim que pot tenir aquesta distància. Aquest valor ha de correspondre a regions de la imatge amb $\rho(\theta)=0$ ó $\rho(\theta)=7\ \forall\theta$ (o sigui, contorns massa proxims al centre de la finestra de transformació o finestres sense contorns).

3.3.3.1. Determinació de la funció distància

Per determinar la funció distància que cal utilitzar s'ha analitzat el comportament de tres funcions distància davant de diferents contorns oberts i tancats. Les funcions distància utilitzades han sigut:

- *Distància euclídea*

  Es tracta de la funció distància més utilitzada per estimar la diferència entre dos vectors. En el nostre cas tractaren la transformada polar del contorn de la regió mostrejada , $\rho(\theta)$, com un vector caracteristic que cal comparar amb el vector patró $\rho(\theta)=3,5\ \forall\theta$ corresponent a una circumferència ideal de radi igual a 3,5.

$$D = \sqrt{\sum_{\theta=0}^{7}[\rho(\theta)-3,5]^2} \qquad \text{[Eq. 3.2]}$$

  Amb aquesta funció distància, l'error total messurat és sempre més petit o igual a la suma parcial dels errors dels radis.

- *Distància Manhattan*

  Aquesta funció és una aproximació a la distància euclidea i consisteix en el sumatori del valor absolut de la diferència entre els valors característics de la mostra i el patró:

$$D = \sum_{\theta=0}^{7}|\rho(\theta)-3,5| \qquad \text{[Eq. 3.3]}$$

  L'avantatge respecte la distància Euclídea resideix en el menor nombre de càlculs que s'han d'efectuar, la qual cosa la fa molt util en aplicacions on el temps de processat és critic. L'inconvenient principal consisteix en donar una mesura distorsionada de la distància ja que l'error total és la suma de l'error en cadasqun dels radis.

- *Distància euclídea quadràtica*

  Amb l'intenció de treure el cost afegit de calcular l'arrel quadrada, tot obtenint els mateixos resultats a l'hora de comparar la singularitat de diferents regions de la imatge, pot ser utilitzada la següent messura de distància:

$$D = \sum_{\theta=0}^{7} [\rho(\theta) - 3{,}5]^2 \qquad \text{[Eq. 3.4]}$$

Aquesta funció distància correspon amb el quadrat de la distància euclidea de la mostra amb el patró model, tot i que també pot ser vista com la variança poblacional no normalitzada de les mostres radials respecte a un valor mig dels radis conegut a priori.

Aquesta distància és sempre superior o igual a qualsevol de les altres.

3.3.3.2. Resultats amb contorns tancats

Es presenten dos resultats diferents amb contorns tancats. En primer lloc els contorns tancats han sigut modelitzats com a circumferències de diferents radis. Com a paradigma de fiabilitat en el reconeixement hem agafat una circumferència de radi igual a 3,5 pixels. Per aquesta circumferència totes les mesures de distància donen el valor zero (fiabilitat = 100%). Per circumferències de radi igual a 0 ó 7 (valors mínim i màxim dels radis repectivament) les funcions distància donen el seu valor màxim (fiabilitat = 0%).

A la figura 3.14. es poden comparar els resultats. Les circumferències resulten més fiables amb una comparació feta amb les distàncies euclídea o el seu quadrat. Amb aquestes distàncies la fiabilitat d'una circumferència de radi diferent a 0 ó 7 és superior al 50% (80% pels radis 2,3,4 i 5). Utilitzant aquestes distàncies reduïnt l'efecte de l'escala sobre la mesura de la fiabilitat.

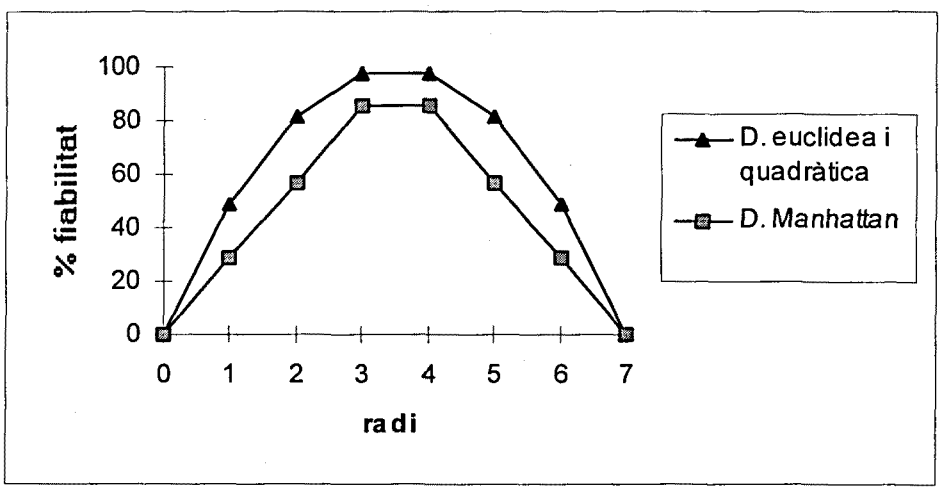


Figura 3.14. Percentatge de fiabilitat en el reconeixement d'una circumferència en funció de la longitud del radi.

En segon lloc, s'han comparat una sèrie de formes diferents amb el patró ideal i s'han obtingut els resultats presentats a la següent gràfica. La geometría de les formes emprades i la seva transformació polar es poden veure a l'annex 1.

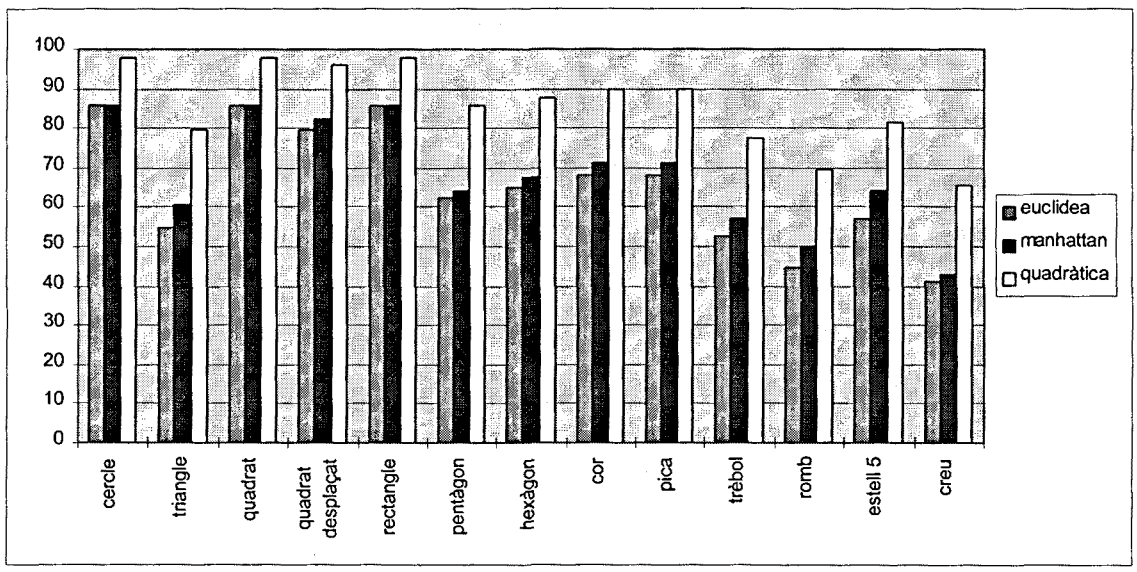


Figura 3.15. Percentatge de fiabilitat en la selecció de diferents formes geomètriques utilitzant les distàncies euclídea, manhattan i quadràtica.

3.3.3.3. Resultats amb contorns oberts

Els contorns oberts presenten en general una fiabilitat més baixa que els contorns tancats degut a que tenen alguns radis amb valors igual a set (saturats). Tot i això la funció de fiabilitat presentada ha donat resultats força significatius en les diferents proves realitzades amb els diferents criteris de distància. Per presentar resultats d'aquestes proves els contorns oberts han sigut modelitzats com un vèrtex d'angle variable (α entre 10 i 180 graus). La fiabilitat ha sigut mesurada a diferents distàncies d'aquests vèrtexs respecte el centre de la finestra de trasformació (*d* de 1 a 7 pixels).

La fiabilitat està indicada com un percentatge respecte la fiabilitat màxima (100%), aconseguida en el cas ideal d'una circumferència de radi igual a 3,5 pixels. (Figures 3.16. a 3.21)

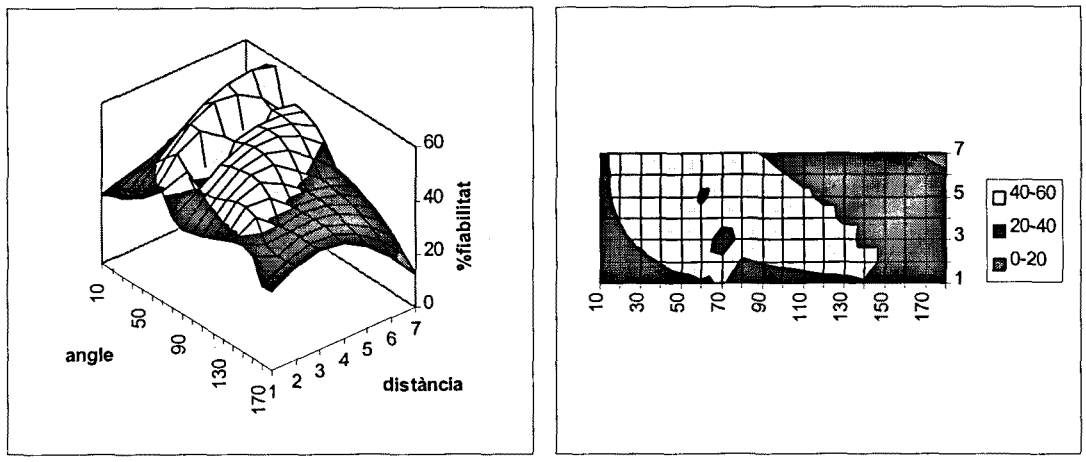


Figura 3.16. % Fiabilitat de regions amb contorns oberts (distància euclídea)

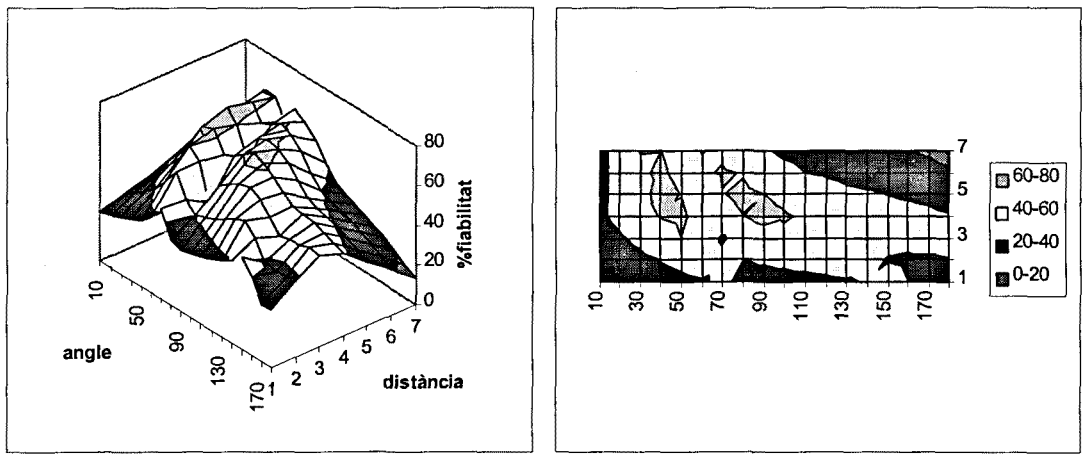


Figura 3.17. % Fiabilitat de regions amb contorns oberts (distància Manhattan)

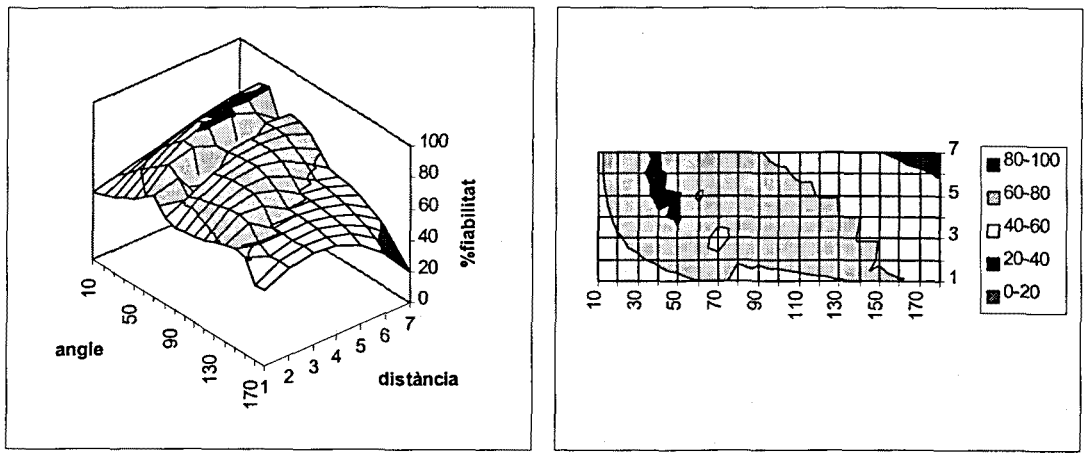


Figura 3.18. % Fiabilitat de regions amb contorns oberts (distància euclídea quadràtica).

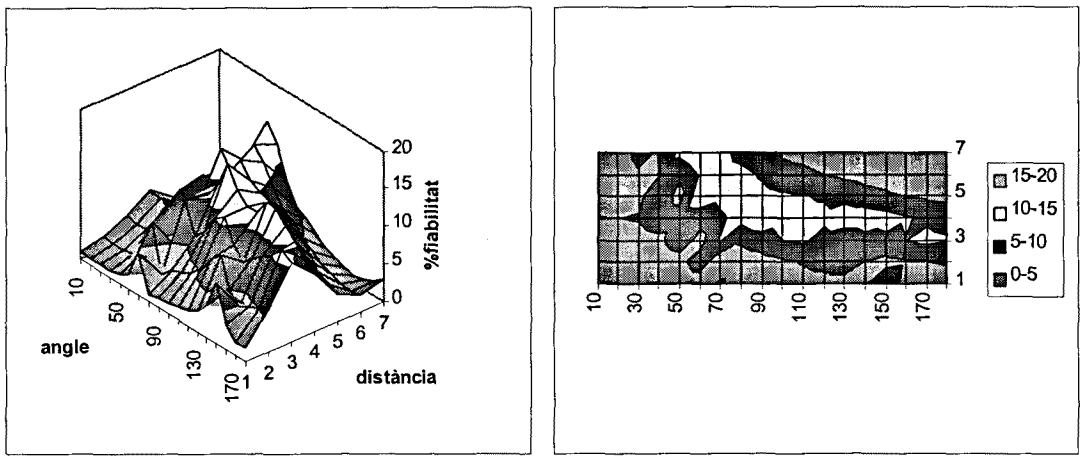


Figura 3.19. Diferència entre Distància Manhattan i Distància Euclídea

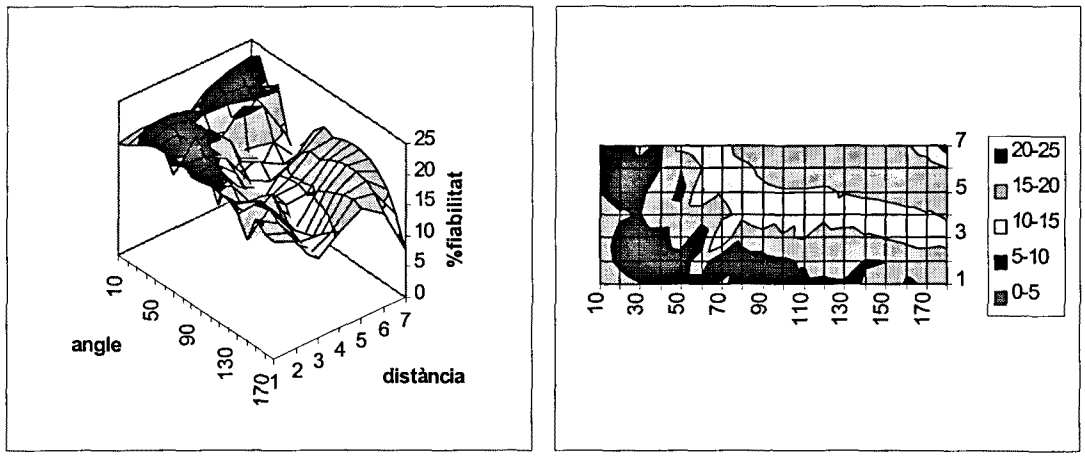


Figura 3.20. Diferència entre Distància Euclídea quadràtica i Distància Manhattan

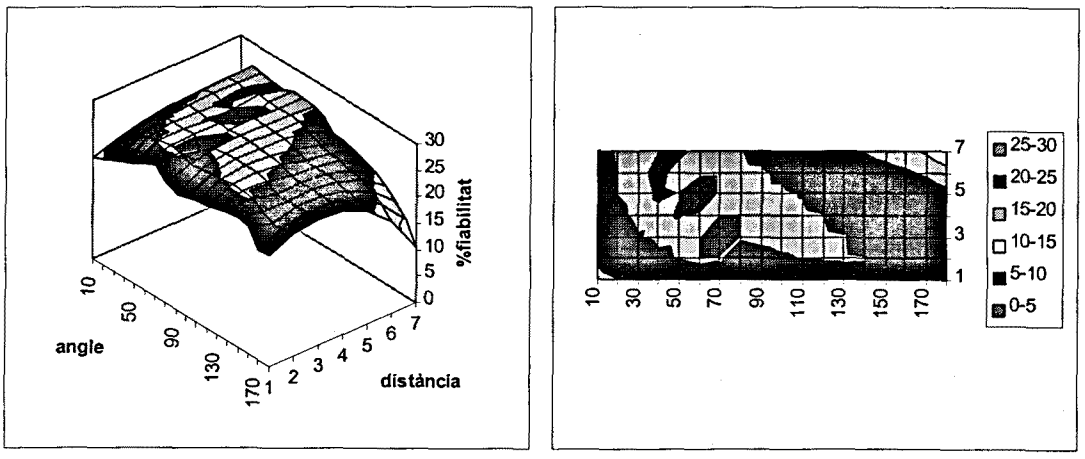


Figura 3.21. Diferència entre Distància Euclídea quadràtica i Distància Euclídea

La mesura de fiabilitat d'una determinada regió de la imatge ha de tolerar canvis petits respecte a la circumferència de radi igual a 3,5 pixels, o bé canvis que afectin de forma semblant als diferents radis. És a dir, ha de tolerar una distribució equitativa de l'error respecte de les components del vector patró. Un cas així es dona per un petit canvi d'escala (apropament/allunyament de l'objectiu respecte les càmeres) o un petit canvi de perspectiva.

Per contra, la funció fiabilitat ha de penalitzar una diferència amb el patró que sigui gran i produïda per un sol radi. Degut a la baixa resolució que té la transformació polar, una regió que resulti amb un radi molt diferent de la resta serà difícil de reconeixer si gira durant el seu seguiment al llarg de la seqüència d'imatges. Aquest era el criteri que es volia maximitzar en l'anàlisi de la homogeneïtat en el valor dels radis (apartat 3.2).

De les funcions distància provades, la que millor sintonitza amb els criteris anteriors és la distància euclídea quadràtica. A més a més, i tal com es pot observar a la figura 3.18, és la que dona fiabilitats més elevades i més homogenies per un conjunt més gran de vertexs.

Els valors obtinguts utilitzant aquestes funcions distància verifiquen que:

$$\textbf{Euclidea} \leq \textbf{Manhattan} \leq \textbf{Euclidea al Quadrat}$$

$$\sqrt{N}*\Delta \leq N*\Delta \leq N*\Delta^2$$

on Δ és la diferència existent en un conjunt de N radis.

O sigui que la distància euclidea al quadrat és la que dona uns valors més elevats de la funció fiabilitat proposada i una major variabilitat (quadràtica).

3.3.4. Efectes de la discretització en la codificació dels radis

La transformada polar de la regió de la imatge presenta una discretització en els possibles valors que poden agafar els radis. Aquesta discretització es deguda a la propia discretització de la imatge en pixels.

El resultat és una discretització també en els valors de la fiabilitat que pot tenir una determinada regió de la imatge tal com es mostra a les figures 3.22 i 3.23.

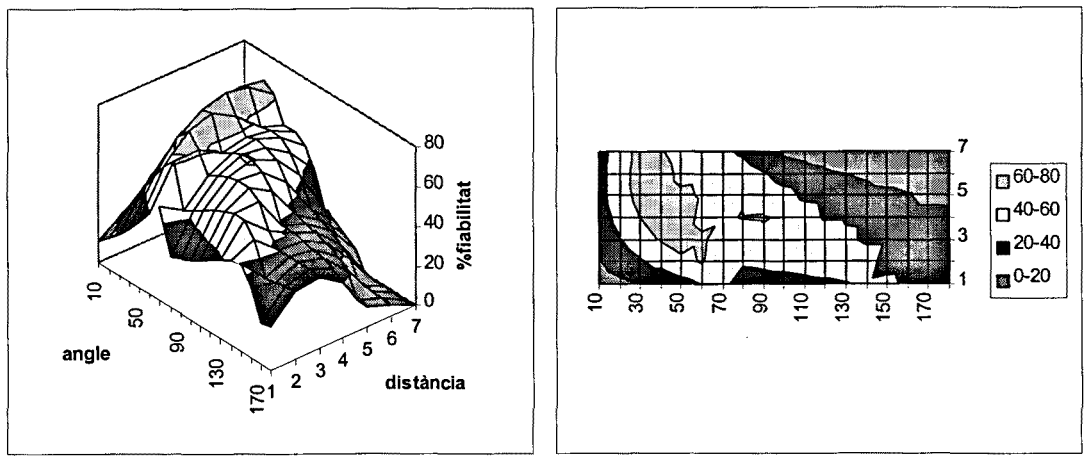


Figura 3.22. % Fiabilitat de regions amb contorns oberts (continua)

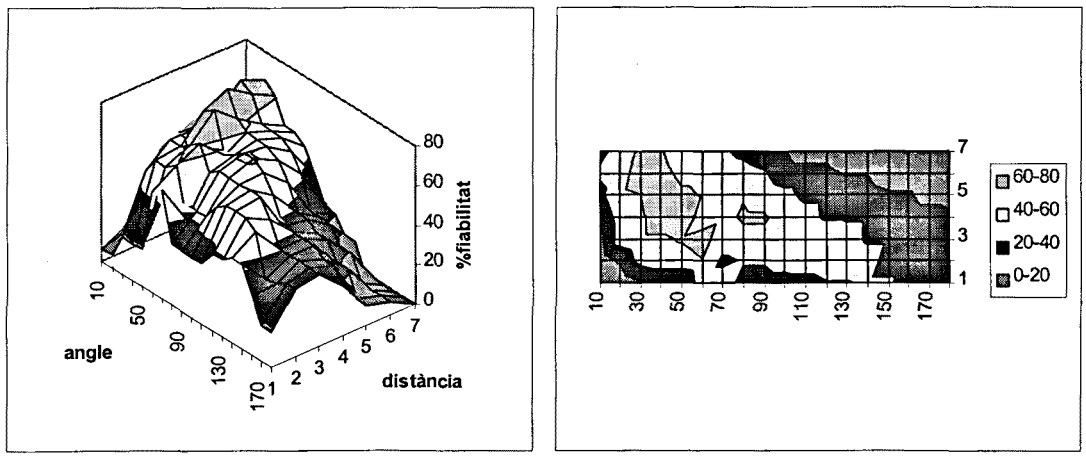


Figura 3.23. % Fiabilitat de regions amb contorns oberts (discreta)

Com es pot apreciar a la gràfica els vèrtexs més fiables són aquells compresos entre els 30° i els 50° a una distància d'entre 4 i 7 pixels respecte al centre de la finestra de transformació. Aquest vèrtexs poden ser reconeguts amb una fiabilitat superior al 60%. Tot i així, els contorns tancats ofereixen una fiabilitat més elevada com pot ser deduït a partir de les figures anteriors. S'ha de tenir en compte que una fiabilitat del 100% s'obté per una circumferència de radi igual a 3,5 pixels (contorn tancat ideal fet servir com a patró).

Com es pot observar a la figura 3.23, l'efecte de la discretització no és critic a l'hora de determinar les regions més caracteristiques de l'escena. Per un estudi més exhaustiu sobre l'incidència de l'error de discretització vegeu el capítol 5 d'aquesta tesi.

Per visualitzar el resultat d'aplicar sobre una imatge aquest criteri de fiabilitat, s'ha presentat el valor de la funció fiabilitat de cada punt de la imatge com un nivell de gris. Aquest valor depèn de la distribució dels pixels de contorn al voltant d'aquest

punt dins de la finestra de transformació (on el punt en qüestió ocupa el pixel central). (Figures 3.24 a 3.26.)

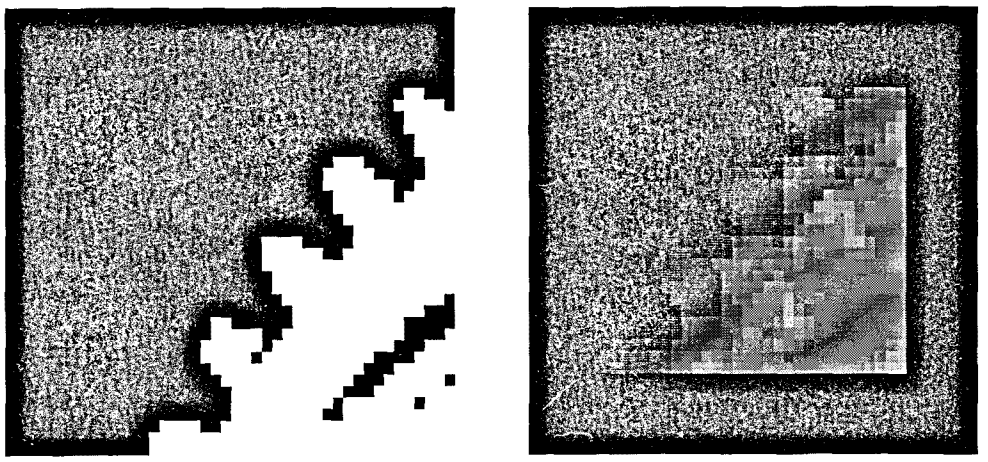


Figura 3.24. Funció fiabilitat de cada pixel d'una regió de la imatge binaritzada a la que es pot veure la fiabilitat relativa dels seus punts. (Recordem que la transformació radial retorna per a cada radi el valor de distància del pixel blanc més proper al centre).

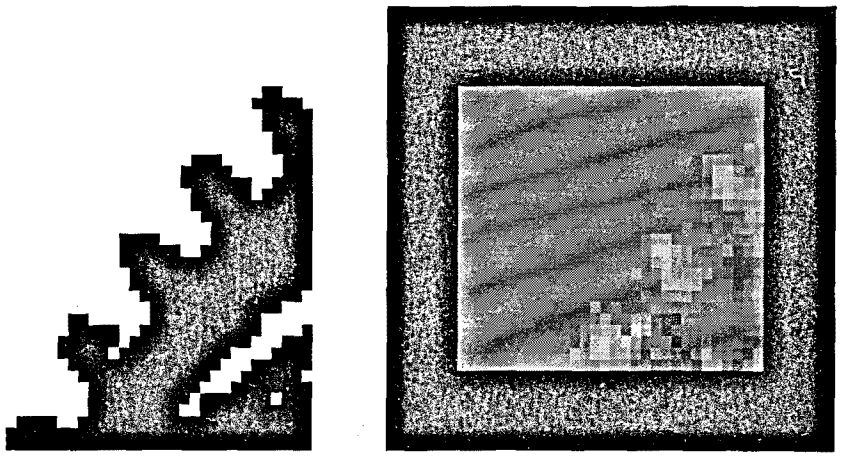


Figura 3.25. La mateixa imatge però amb el binaritzat simètric. (Es pot observar com han canviat els punts caracteristics trobats per la funció fiabilitat).

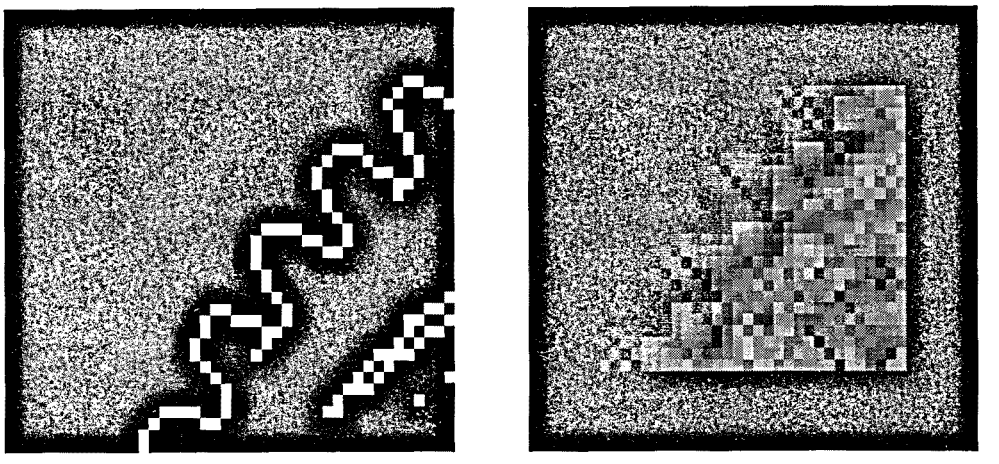


Figura 3.26. Amb l'imatge contorn es pot veure com es complementen els punts característics trobats a les dues imatges anteriors.

Els resultats d'aplicar aquesta funció fiabilitat a un ampli conjunt d'imatges demostren que la funció proposada es capaç de discriminar aquelles formes que permeten un millor reconeixement i localització. Tot i així es pot observar a la gràfica de la figura 3.23, que un contorn recte (vertex de 180 graus) a una distància de 2 o 3 pixels respecte al centre de la finestra de transformació dona fiabilitats en el seguiment de gairebé el 40%. Un contorn recte no pot ser localitzat en les dues dimensions del pla, ja que únicament permet la seva localització en la direcció perpendicular al propi contorn. Una linia recta per tant, hauria de donar una fiabilitat molt més baixa.

Això implica utilitzar una sensibilitat molt elevada per a determinar el llindar de decisió, per la qual cosa s'ha imposat una selecció prèvia del tipus de característica local que busquem: aquelles que tenen una transformació polar que permet la seva localització precissa en les dues dimensions del pla de la imatge.

3.3.5. Detecció de vèrtex.

Amb els condicionaments desfavorables que presenta (en alguns casos concrets) la funció de selecció proposada, la solució adoptada ha sigut descartar de forma explícita totes aquelles regions de la imatge que no tinguin un contorn que talli com a mínim 5 radis consecutius.

Aquest criteri és equivalent a demanar al contorn una certa curvatura (en funció de la distància al centre) ja que una recta només pot tallar com a molt 4 radis consecutius (figura 3.27.). Per una finestra de transformació de 15x15 pixels, el radi màxim de la corba que pot passar aquest filtre és de 13 pixels.

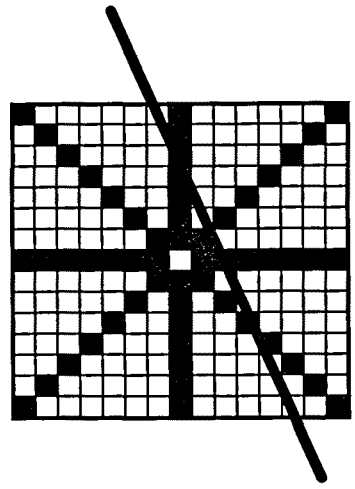


Figura 3.27. Una recta només pot arribar a tallar 4 radis consecutius

El criteri de continuïtat és necessari per poder eliminar aquelles combinacions no desitjades de punts de contorn corresponents a contorns de diferents objectes (contorns no continus) que apareixen sovint dins de la finestra de transformació degut a la proximitat entre els diferents objectes a la imatge (figura 3.28.)

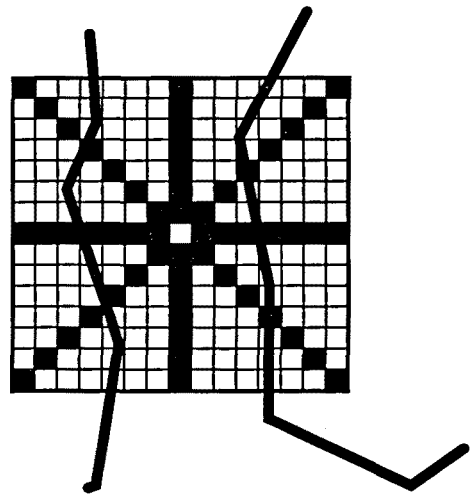


Figura 3.28. Exemple de contorns no continus (dins de la finestra de transformació) que donen una fiabilitat elevada amb un criteri que no contempli restriccions de continuitat.

Contorns tancats prou petits compleixen el doble criteri de continuitat en el contorn i de curvatura mínima (habitualment els 8 radis estan tallats pel mateix contorn). Llavors el criteri de fiabilitat amb restriccions de continuïtat, segueix sent vàlid per aquests contorns.

Pel que fa als **contorns oberts**, modelitzats com a vèrtexs a l'apartat 3.3.2, la funció fiabilitat obtinguda dona els resultats de la figura 3.29 (convé comparar-la amb la figura 3.23):

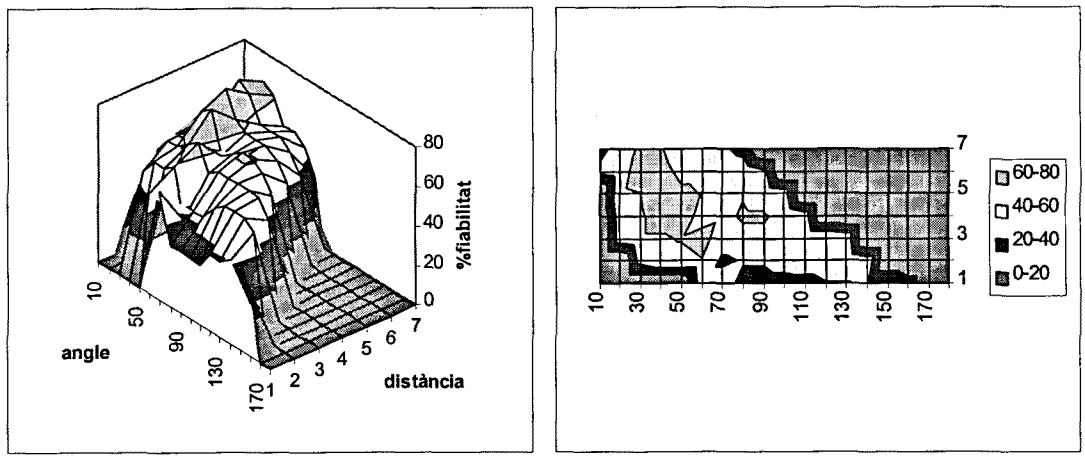


Figura 3.29. Fiabilitat de regions amb contorns oberts (discreta) amb la restricció de tallar més de quatre radis consecutius.

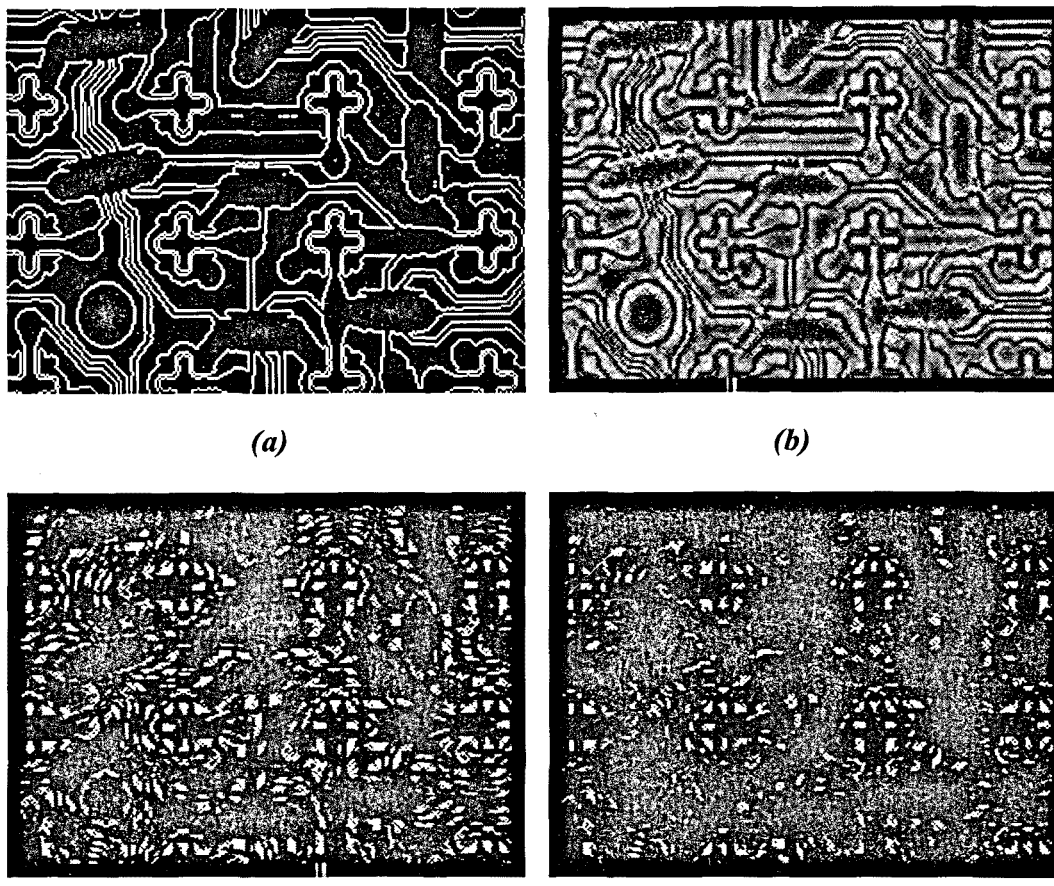
El resultat d'efectuar aquest filtrat ha sigut l'anul·lació de tots aquells vèrtexs que presenten una combinació angle-distància (α,d) representada amb color negre a la figura 3.29.

Aquests vèrtexs eliminats són:

- Els vèrtexs d'angle inferior a 18°.
- Els vèrtexs d'angle inferior a 57° i que es trobin a una distància d'un pixel respecte al centre de la finestra de transformació.
- Els vèrtexs amb combinacions (α,d) situades per sobre i a la dreta de la recta (75°,7) a (150°,1). Corresponents a vèrtexs oberts i/o lluny del centre de la finestra de transformació.

Qualsevol d'aquests supòsits dificulten enormement la identificació i seguiment d'un vèrtex, ja que es poden confondre fàcilment amb una línia recta i per tant es fa difícil la seva localització precisa en les dues dimensions del pla de la imatge.

El resultat d'aplicar la funció fiabilitat definida amb el filtre de continuïtat sobre la transformació polar de cada pixel d'una imatge es pot veure a la figura 3.30. Un estudi més extés, amb resultats per diferents dimensions en la finestra de transformació pot ser trobat a l'annex 3 d'aquesta tesi.



(a) *(b)*

(c) *(d)*

Figura 3.30. Imatges d'un circuit imprès:

(a) imatge de contorns. ***(b) fiabilitat de cada posició sense filtres.***
(c) filtre de 4 radis consecutius. ***(d) filtre de 5 radis consecutius***

Com es pot apreciar a les imatges de la figura 3.30, l'imposició introduïda de tallar quatre radis consecutius de la transformació polar d'una regió, no descarta de forma definitiva la selecció de contorns molt oberts (veure la figura 3.30.(c)). Es fa per tant necessària la condició imposada de cinc radis consecutius per aconseguir una bona selecció de característiques locals.

En la implementació final d'aquest procés de selecció automàtica de característiques locals, aquest filtratge de continuïtat s'efectua de forma prèvia a l'estudi de fiabilitat amb la aplicació de la funció distància. Es fa en aquest ordre, degut a que per tractar-se d'un procés ràpid de filtratge, permet un temps d'inicialització del sistema més reduït.

El mòdul d'inicialització escolleix com a característiques locals per efectuar l'aparellament i posterior seguiment, aquelles que donen un valor més elevat. La selecció pot venir donada per un llindar d'aceptació o bé per un determinat nombre de característiques demanades per l'usuari. Cal dir però, que independenment del valor de la fiabilitat, només s'agafen com a característiques locals punts que no estiguin en contacte dins de la imatge, ja que aquests punts pertanyen a la mateixa característica local.

La informació obtinguda a partir de la imatge en aquesta etapa de selecció queda llavors reduïda a una llista de característiques locals amb la següent informació:

- posició dins de la imatge (x,y)
- vector característic consistent en la transformada polar discreta de la regió centrada al punt (x,y).

Aquesta informació és utilitzada per efectuar l'aparellament estèreo.

3.4. Aparellament estereoscòpic

Un mètode habitual per obtenir informació relativa a la distància a la que es troben els objectes presents en l'escena és la estereoscopia, que consisteix en l'adquisició d'una parella d'imatges mitjançant dues cameres desplaçades lateralment l'una de l'altra una distància coneguda. Després, cal identificar aquells punts de les imatges esquerra i dreta que es corresponen amb el mateix punt de l'escena (coneguts com a *punts homòlegs o "conjugate pair"*). Una vegada identificats, i coneguda la geometria del sistema, es pot establir la seva distància respecte les cameres. Detectar aquestes parelles de punts entre imatges és però, un problema mal condicionat de difícil solució conegut com el *problema de la correspondència.*

En el present apartat es proposa una metodologia per abordar aquest problema, basat en un reconeixement local de la imatge, tot utilitzant la transformació polar de la imatge descrita amb anterioritat en aquesta tesi. Es suposarà que s'ha fet una selecció prèvia sobre una de les imatges (concretament la imatge esquerra) dels punts candidats a ser aparellats, abans del procés d'aparellament dels punts homòlegs a l'altre imatge. Això reduirà per una part el problema de la correspondència i per l'altra l'alt cost computacional que té l'aparellament.

3.4.1. Geometria del sistema estereoscòpic

A la figura 3.31 es mostra la geometria del sistema estereoscòpic utilitzat en aquesta tesi i que es correspon amb el model estereoscòpic anomenat *"pin-hole"*.

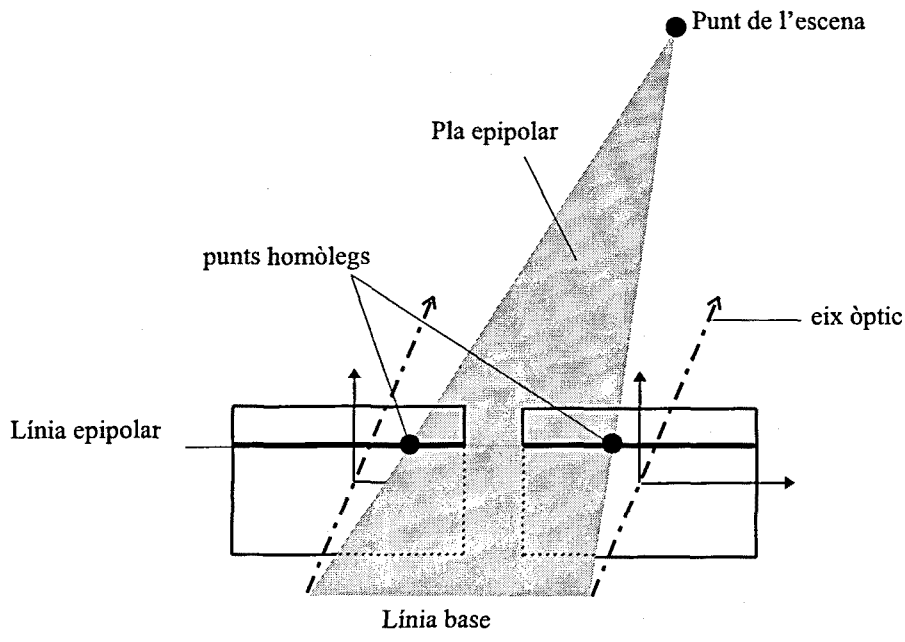


Figura 3.31. Geometria del sistema estereoscòpic

El sistema consisteix en la utilització de dues cameres idèntiques separades únicament en la direcció x per una distància igual a b (anomenada línia base o *"baseline"*). Els plans d'imatge de ambdues cameres són coplanars en aquest model. Un determinat punt de l'escena és vist per les dues cameres en diferents posicions als respectius plans d'imatge. El desplaçament existent entre les posicions de les dues projeccions del punt en qüestió a les imatges esquerra i dreta s'anomena *disparitat*. Aquesta disparitat es pot mesurar sempre al llarg d'una recta paral·lela a l'eix de les x anomenada *línia epipolar*.

S'anomena *pla epipolar* al pla definit per la línia base i el punt de l'escena. La intersecció del pla epipolar amb els plans d'imatge (coplanars) defineix la *línia epipolar*.

Pel model estereoscòpic que mostra la figura 3.31, els punts projectats sobre la imatge esquerra resulten sempre projectats a la mateixa línia de la imatge dreta. Es a dir, la línia epipolar coincideix amb una línia del pla d'imatge, el que facilita la búsqueda dels punts homòlegs dins de la imatge degut a que ara la disparitat només pot ser horitzontal. La majoria d'autors utilitzen també aquest model de projecció després d'efectuar un calibratge previ de les cameres [Faugueras,92] [Jones,92] [Papadimitriou,96].

En la pràctica, però, sempre existeix una certa disparitat vertical que pot variar de ±1 a ±3 pixels i que pot variar dins de la imatge. Això és degut, entre d'altres factors, a una falta de cal.libració entre les cameres (orientació relativa diferent de zero), a la diferent distorsió entre les cameres que no són idèntiques i al mateix procés de discretització [Olsen, 92].

3.4.2. Càlcul de la profunditat

La informació coneguda a priori sobre la geometria del sistema servirà per a fer el càlcul de la distància a la que es troben els punts de l'escena.

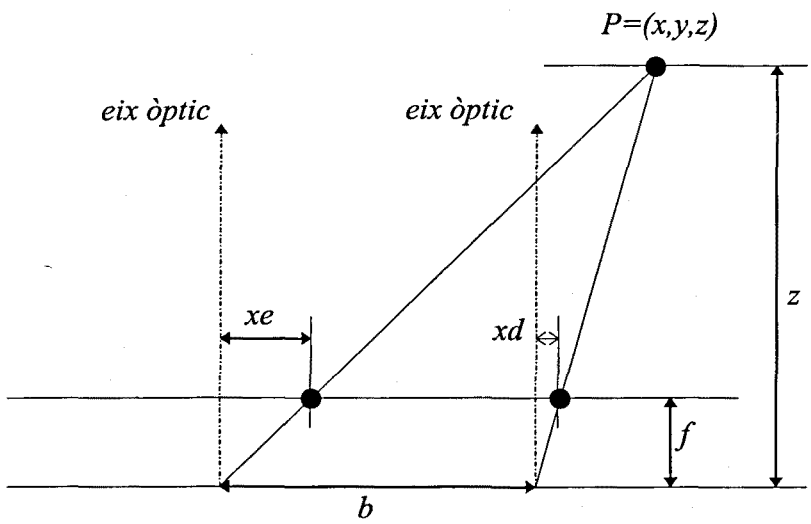


Figura 3.32. Modelització matemàtica del sistema estereoscòpic

A la figura 3.32 es mostra la modelització matemàtica del sistema, en que el punt P és observat als punts xe i xd als plans d'imatge esquerrà i dret, respectivament. Suposem que l'origen del sistemes de coordenades del món és a la meitat de la línia que separa les dues cameres (baseline). De fet la posició de l'origen del sistema de coordenades del món al llarg d'aquesta línia és totalment arbitrària ja que no afecta al càlcul de distància z del punt P. D'aquesta forma obtenim les següents identitats entre triangles, formats per la línia Pper la camera esquerra:

$$\frac{x+\frac{b}{2}}{z}=\frac{xe}{f}$$

i per la camera dreta:

$$\frac{x-\frac{b}{2}}{z}=\frac{xd}{f}$$

restant les dues equacions, obtenim:

$$\frac{b}{z}=\frac{(xe-xd)}{f}$$

de forma que queda:

$$z=\frac{b.f}{(xe-xd)} \qquad \text{[Eq. 3.5]}$$

El que ens diu que la profunditat dels punts de l'escena és donada per l'invers de la distància entre els punts homòlegs al pla de la imatge (disparitat) i la constant $b.f$.

Cal remarcar algunes dades relatives a la disparitat. Degut a la geometria del sistema, es verifica sempre que $xe \geq xd$, la qual cosa garanteix que la disparitat sempre és positiva i per tant també ho és la z. Degut al càlcul mitjançant l'invers de la disparitat el comportament de z no és pas lineal i els seu valor es veu molt afectat per l'error de discretització quan la disparitat és petita. (Per un estudi més detallat veure el capítol 5 d'aquesta tesi dedicat a l'anàlisi d'errors)

Si es vol augmentar la precisió de la mesura z, amb uns paràmetres de camera determinats, podem fer-ho augmentant la distància que separa les cameres, és a dir b. D'aquesta forma, creix la disparitat, i l'error degut a la discretització té menys influència en el càlcul de z (per una determinada distància).

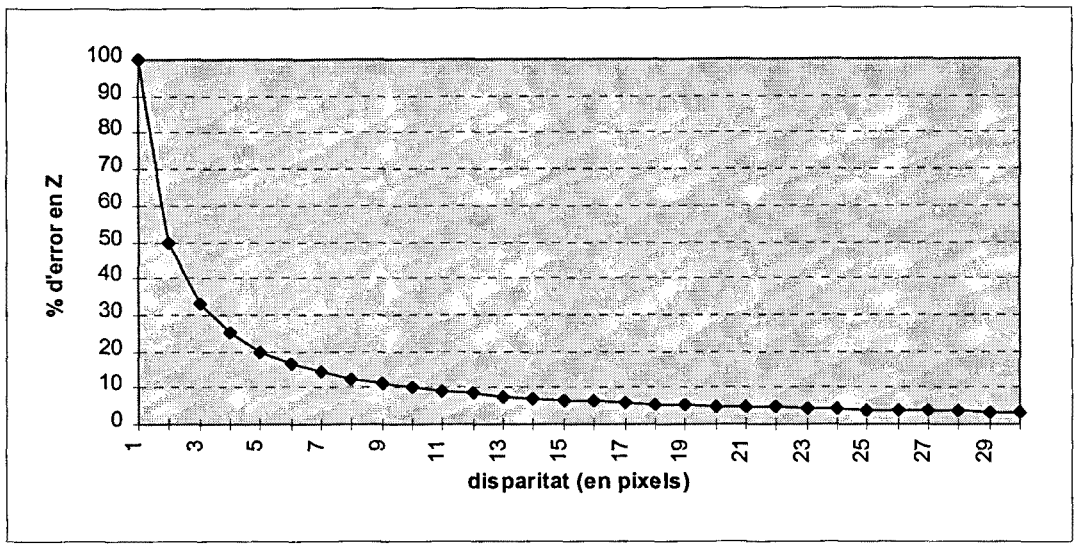


Figura 3.33. Percentatge d'error que pot causar la discretització de la imatge en la mesura z en funció de la disparitat (en pixels).

D'altra banda, augmentar la separació entre les cameres comporta altres problemes. Per exemple, quan la distància *b* augmenta, la fracció de imatge que pot ser vista per ambdues cameres decreix. També augmenten les oclusions provocades pels propis objectes de l'escena, el que comporta l'aparició de "*zones mortes*" sense informació de profunditat. A més a més, les parts de l'escena que poden ser vistes per les dues cameres es projecten de una forma més diferent degut a la distorsió introduïda per la perspectiva, fent més difícil l'aparellament dels punts homòlegs.

3.4.3. Localització dels punts homòlegs

El problema de la correspondència pot ser enunciat com*: per cada punt a la imatge esquerra, trobar el seu homòleg a la imatge dreta.* Per determinar la correspondència entre punts és fa necessària una *mesura de similitud* entre els punts.

És evident que el punt a ser aparellat ha de tenir característiques clarament diferents de la resta de punts del seu voltant, ja que d'una altra forma seria difícil identificar-ho respecte els punts veïns. El problema de la correspondència comença llavors amb la selecció dels punts a ser aparellats i les característiques que els defineixen.

La diferència fonamental entre els mètodes d'aparellament estereoscòpic radica en la forma de mesurar la similitud entre els punts de les dues imatges, diferenciant-se els que utilitzen *característiques relatives als contorns* dels que usen *característiques relatives a regions de la imatge.* Una relació més completa dels mètodes ja ha sigut presentada al capítol 1.

La idea bàsica en els algorismes d'aparellament consisteix en primer lloc en la selecció dels punts candidats de la imatge esquerra i dreta. Posteriorment, cada punt

seleccionat en la imatge esquerra és comparat amb aquells de la imatge dreta que cauen dins de la línia epipolar. Hom suposa que el procés de selecció es pot fer en paral·lel en les dues imatges, ja que d'altra forma es fa més lent l'aparellament.

A la solució proposada els punts són seleccionats només a partir d'una de les imatges (l'esquerra concretament). Posteriorment, es busquen els seus homòlegs dins de l'altra imatge (dreta), limitant la búsqueda a dins de la línia epipolar, que per cal.libració prèvia de les cameres seran coincidents amb les pròpies línies de la imatge amb un petit error tolerat pel propi sistema tal i com mostren els resultats obtinguts (figura 3.39).

La raó de fer-ho així, no és una altra que estalviar temps de càlcul en la inicialització del seguiment tridimensional. Efectivament, si el nombre de característiques locals seleccionades és relativament petit, només cal buscar l'aparellament dins les línies epipolars en la imatge dreta, el que resulta menys costós que una selecció de punts en la globalitat de la imatge.

A més a més, amb aquest sistema la garantia de coincidència del punt trobat amb el punt homòleg real és més elevada que en el cas de la selecció prèvia a l'aparellament. Això és degut a que els criteris de selecció (basats en un *factor de singularitat*) no garanteixen la selecció dels dos punts homòlegs a la imatge esquerra i dreta degut a problemes de perspectiva, il·luminació i diferència de captació entre les cameres, entre d'altres.

El mètode proposat en aquesta tesi per solucionar el problema de la correspondència està basat en el *reconeixement local* de la imatge. Per fer aquest reconeixement s'utilitza com a vector característic la transformació polar de la regió de la imatge que conté el punt en qüestió. Aquesta transformació està basada en la descripció polar d'un contorn i a sigut optimitzada tal i com va ser exposat a l'inici d'aquest capítol.

La solució presentada és, per tant, un híbrid entre les dues categories en que s'acostumen a classificar els mètodes d'aparellament, ja que, per fer la mesura de similitud entre punts es fa servir la transformació polar d'una regió de la imatge en la que prèviament s'han extret els contorns.

El mètode de reconeixement està basat en una mesura de distància entre el vector característic de la regió de la imatge esquerra i els vectors característics de les regions de la imatge dreta centrades a la mateixa línia epipolar. La funció distància utilitzada és una generalització de l'obtinguda a l'apartat 3.3 d'aquest capítol (Eq. 3.4) per a la selecció de les característiques locals. D'aquesta forma es defineix:

$$D(j) = \sum_{\theta=0}^{7} \left[r(\theta) - m_j(\theta) \right]^2 \qquad \forall j \in \textit{línia epipolar} \qquad [\text{Eq. } 3.6]$$

On $D(j)$ és una funció discreta definida al llarg de la línia epipolar, els valors de la qual coincideixen amb la distància (falta de similitud) del vector característic de la regió de la imatge esquerra $r(\theta)$ i els vectors característics associats a cada pixel dins de la línia epipolar $m_j(\theta)$.

De fet *D(j)* només cal que sigui calculada dins d'un interval de la línia epipolar. Aquest *interval de búsqueda* vindrà determinat per una disparitat màxima i una disparitat mínima. Aquestes disparitats han de ser donades a priori en funció del rang de distàncies esperades dels objectius que es volen seguir i que poden canviar amb l'aplicació concreta que es vulgui fer del sistema.

Habitualment s'acostuma a definir la disparitat mínima com a zero (z = ∞), fixant-se només la disparitat màxima (z mínima).

Suposant que la característica local que volem reconèixer ocupa la posició *i* a l'eix de les x de la imatge esquerra, es redefineix *D(j)* com:

$$D(j) = \sum_{\theta=0}^{7} \left[r(\theta) - m_j(\theta) \right]^2 \qquad \forall j \in [i\text{-}disparitat_{max} , i\text{-}disparitat_{min}]$$

Una vegada avaluada la funció *D(j)* al llarg de l'interval de búsqueda (Figura 3.34), es passa a seleccionar el millor candidat a punt homòleg d'entre els punts que pertanyen a l'interval.

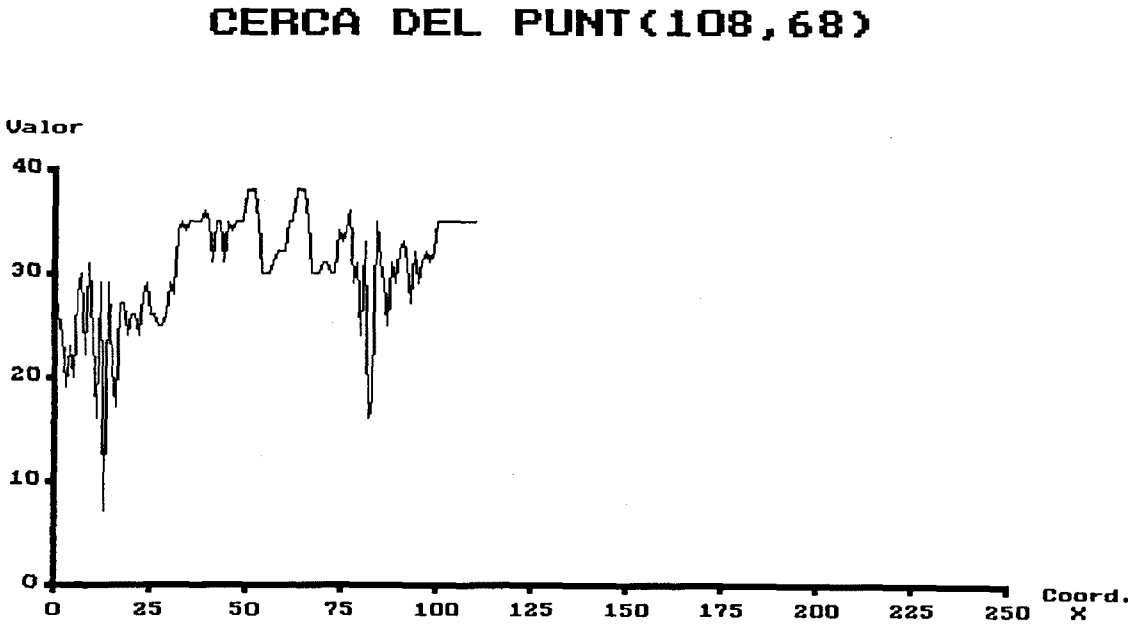


Figura 3.34. Resultat de la funció D(j) (per un determinat cas) dins de l'interval de búsqueda. A les ordenades s'indica la disparitat (en pixels).

Els criteris adoptats per efectuar l'aparellament són:

1. El punt homòleg es trobarà a la posició *j* on *D(j)* es fa mínima dins de l'interval.

2. Que la funció *D(j)* tingui un màxim local al costat del punt homòleg (o als dos costats) a una distància màxima de 7 pixels.

3. Que el valor de la funció *D(j)* sigui inferior a un determinat *llindar d'acceptació de l'aparellament.*

El primer criteri és clarament aquell que es persegueix amb la definició de *D(j).* Òbviament el millor candidat és aquella posició *j* on la funció *D(j)* té el seu mínim dins de l'interval.

El segon i tercer criteris han sigut inclosos a partir dels resultats obtinguts i marquen filtres a la primera condició. Aquests criteris es fan imprescindibles per augmentar la fiabilitat de l'aparellament.

La presència d'un màxim local al costat del mínim de la funció *D(j),* evidencia la proximitat d'un contorn pel que necessàriament la línia epipolar ha de passar en cas de que realment contingui el punt homòleg buscat. Recordem que en la definició de característica local es va incloure la necessitat de que el contorn tallés més de cinc radis consecutius. Aquesta condició treu la possibilitat d'agafar un mínim que es doni en una regió homogènia dins de l'interval seleccionat a la imatge de contorns.

En quan al tercer criteri, bàsicament es tracta de rebutjar a priori aquells aparellaments amb escassa probabilitat d'encert i que amb tota probabilitat causarien incoherències durant la fase de seguiment. En aquest cas el valor de la funció *D(j)* actua com un indicador de la fiabilitat que el punt homòleg es trobi realment a la posició *j* .

3.4.4. Resultats de l'aparellament.

Es mostren a continuació els resultats obtinguts sobre algunes parelles d'imatges a les que s'ha aplicat el mètode d'aparellament de característiques locals proposat. El mètode ha estat provat tant en imatges sintètiques com també en moltes i diverses imatges reals. Per avaluar el mètode s'han tingut en compte dos criteris:

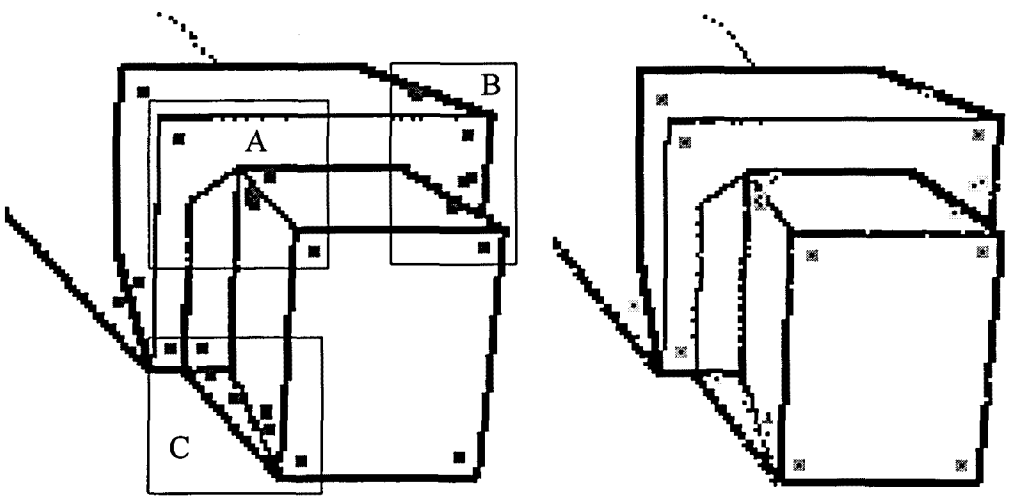
1. El percentatge de característiques locals seleccionades a la imatge esquerra que el sistema ha pogut aparellar (s'han verificat les condicions sobre *D(j)*).

2. El percentatge de característiques locals que són aparellades de forma correcta (respecte a les aparellades).

El percentatge de característiques aparellades és per si sol poc significatiu, ja que pot ser ajustat variant el llindar d'acceptació de l'aparellament (valor màxim permés per *D(j)*). Evidenment a messura que baixem el llindar, baixa el percentatge de característiques aparellades, però, el percentatge d'encerts creix.

El valor més significatiu és el percentatge d'encerts respecte la quantitat d'aparellament efectuats pel sistema. O sigui, quants d'aquest aparellaments han estat correctes. En les proves realitzades, aquesta quantitat, a diferència de l'anterior, no ha estat calculada de forma automàtica. Ha sigut necessària una supervisió dels resultats obtinguts utilitzant el mètode d'aparellament proposat.

El fet de tenir un percentatge d'aparellaments no massa elevat (primer percentatge), no representa cap problema mentre el nombre de característiques seleccionades inicialment a la imatge esquerra sigui prou gran com per garantir un nombre d'aparellament, diguem, 'adequat' a l'aplicació. El que resulta més desfavorable és un error en la selecció de punts homòlegs (segon percentatge) ja que causa un error en la estimació inicial de la profunditat de l'objecte. A partir de les proves realitzades s'ha constatat però, que els errors més habituals en l'aparellament són deguts a la presència d'un error de localització entre les imatges de contorn esquerra i dreta, causant un error en la mesura de la disparitat de 1 ó 2 pixels



Característiques locals seleccionades ***Punts homòlegs trobats***

IMATGE R37

Nombre de característiques locals seleccionades....	27	
Nombre de característiques locals aparellades.....	23	(85%)
Nombre d'aparellaments considerats correctes.....	21	(91%)
Nombre d'aparellaments considerats incorrectes.....	2	(9%)

Figura 3.35. Aparellament de dos blocs situats sobre una taula de treball. El nivell de gris dels punts homòlegs representa el valor de la funció D(j) en aquell punt.

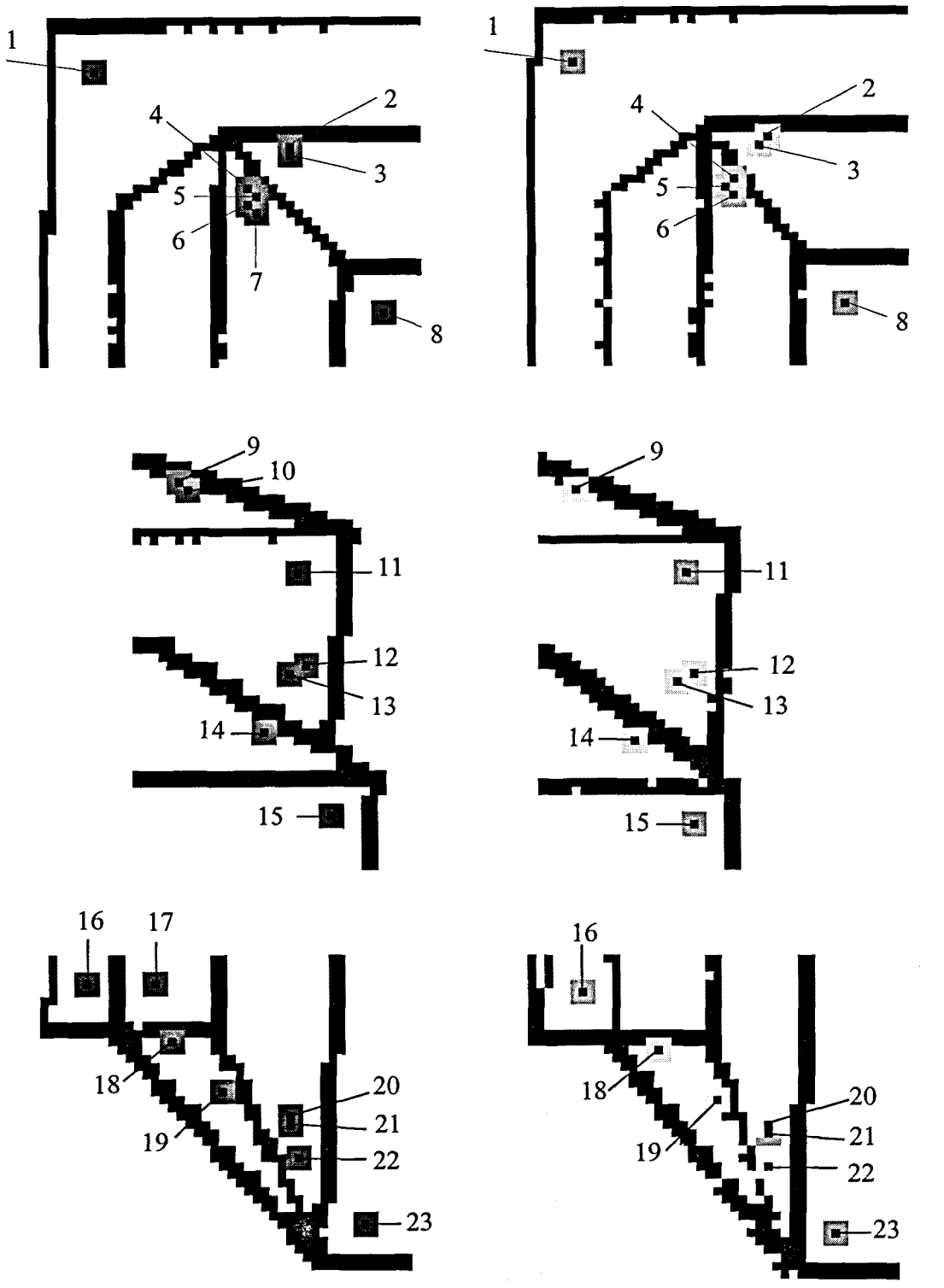


Figura 3.36. Detall dels aparellaments de les zones A, B i C de la imatge. Els punts 7, 10 i 17 no han resultat aparellats. Els punts 3 i 5 presenten un error de localització de 1 i 2 pixels respectivament.

Per imatges sintètiques de formes i contorns ben definits, la fiabilitat del mètode d'aparellament ha resultat gairebé absoluta (amb percentatges d'aparellaments correctes sempre superiors al 95%) per percentatges d'aparellament superiors al 80%.

En el cas d'imatges reals, els resultats no han estat tant sorprenents, degut principalment a problemes amb la il·luminació i textures del objectes, que pot provocar imatges de contorn fortament diferents a les projeccions esquerra i dreta.

En mitjana, per a les imatges reals provades, s'ha obtingut un percentatge d'aparellaments correctes força elevat (superior al 90% en molts casos) a costa de penalitzar el nombre d'aparellaments considerats vàlids a partir de l'anàlisi de la funció $D(j)$ ($\cong$50%). El llindar d'acceptació de $D(j)$ en aquests casos cal que sigui molt baix (<10).

La cal·libració de les càmeres, també ha supossat un obstacle a l'hora d'assegurar que les linies epipolars es corresponguin amb les linies d'imatge. El sistema proposat s'ha mostrat molt robust a errors de fins a ±2 línies en la localització dels punts homòlegs, la qual cosa indica que pot ser utilitzat en sistemes que no puguin garantir una cal·libració precisa durant el seu funcionament (com sería el cas de vehicles mòbils). Una mostra dels resultats obtinguts per la funció $D(j)$ en diferents línies pot ser trobada a la figura 3.39.

En qualsevol cas, i com serà expossat amb detall a l'apartat 3.6.2 , les parelles de punts que no corresponguin amb el mateix punt físic de l'escena provocaran durant el seu seguiment (amb molta probabilitat), una incoherència en les dades relatives a la seva trajectòria, amb la qual cosa el seu seguiment serà abortat.

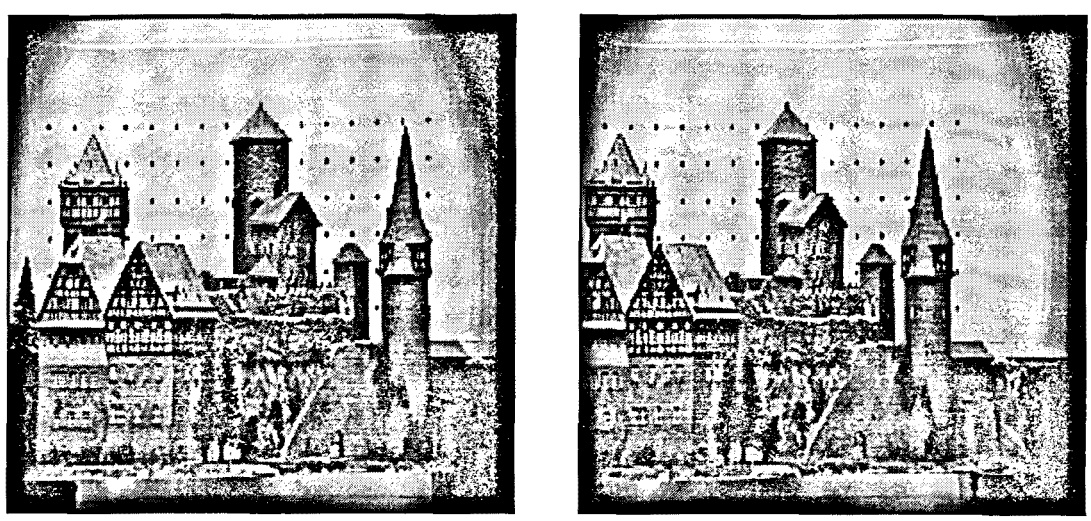


Figura 3.37. Parella de imatges amb elevada textura

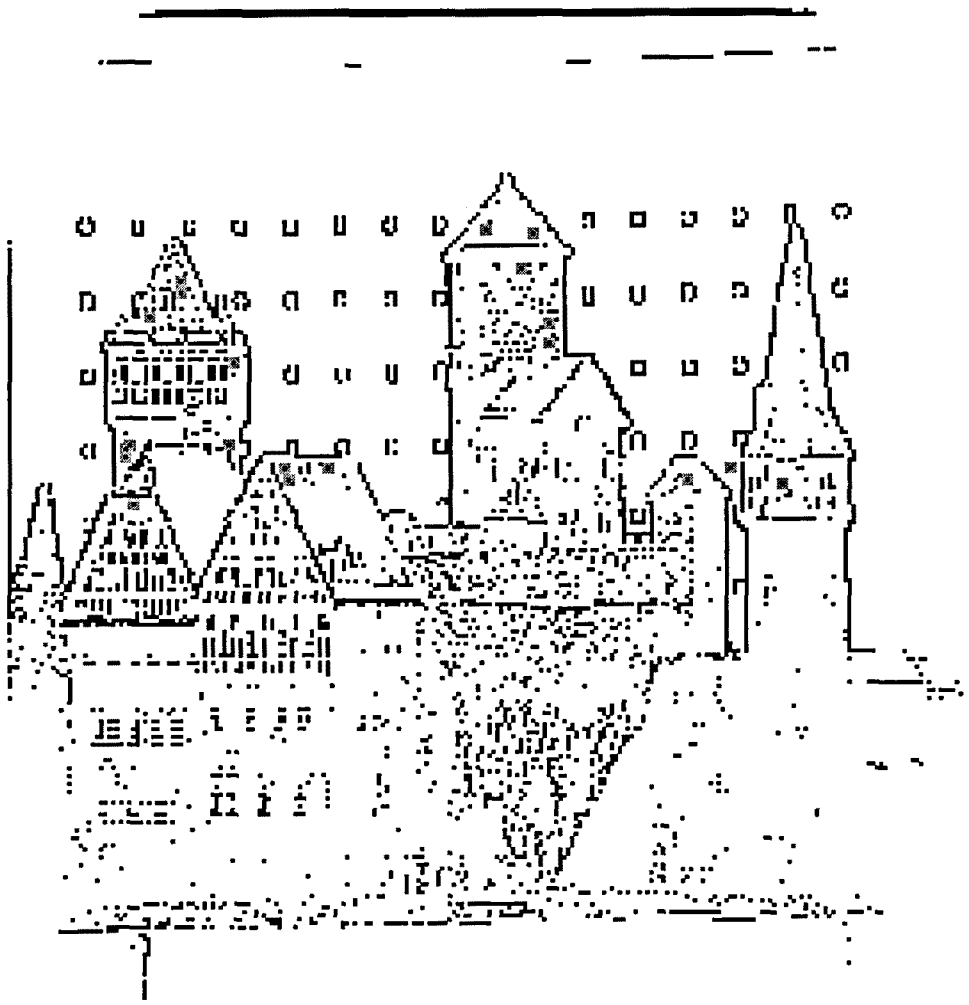


Figura 3.38 (a). Imatge de contorns esquerra.

Característiques locals seleccionades.

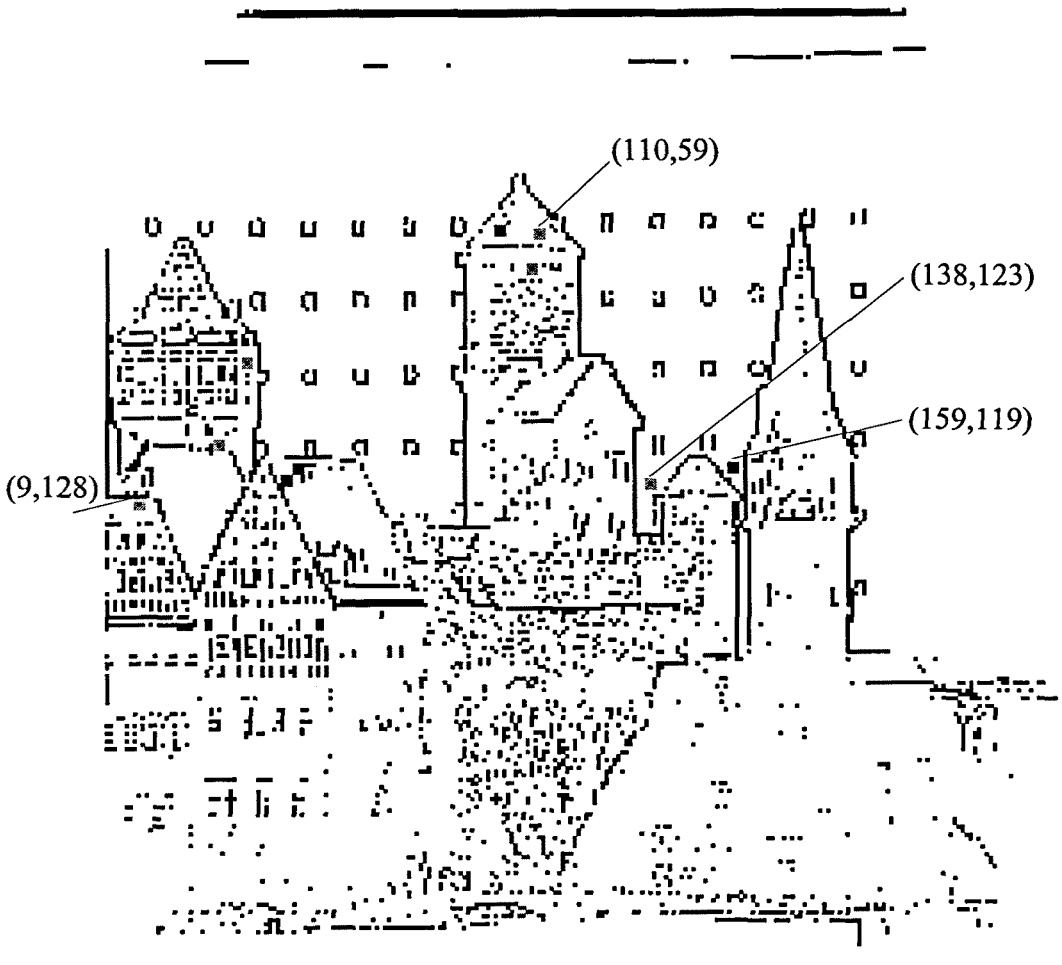


Figura 3.38 (b). Imatge de contorns dreta. Punts homòlegs trobats.

IMATGE R53

Nombre de característiques locals seleccionades....	20	
Nombre de característiques locals aparellades.....	10	(50%)
Nombre d'aparellaments considerats correctes.....	8	(80%)
Nombre d'aparellaments considerats incorrectes.....	2	(20%)

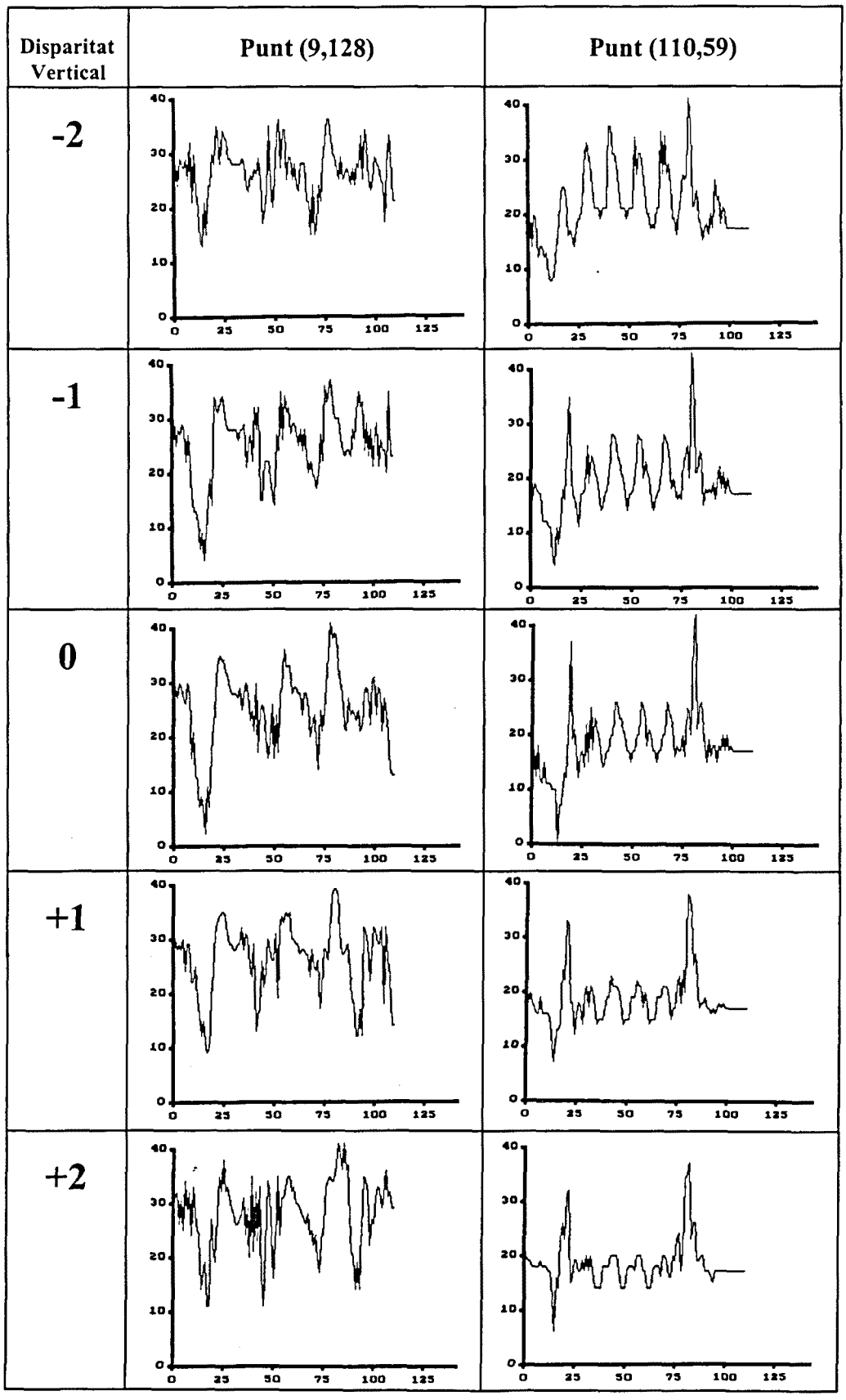


Figura 3.39. (a). Evolució de la funció D(j) per diferents linies d'imatge al voltant de la linia epipolar. Es pot observar el manteniment del punt on D(j) es fa mínima.

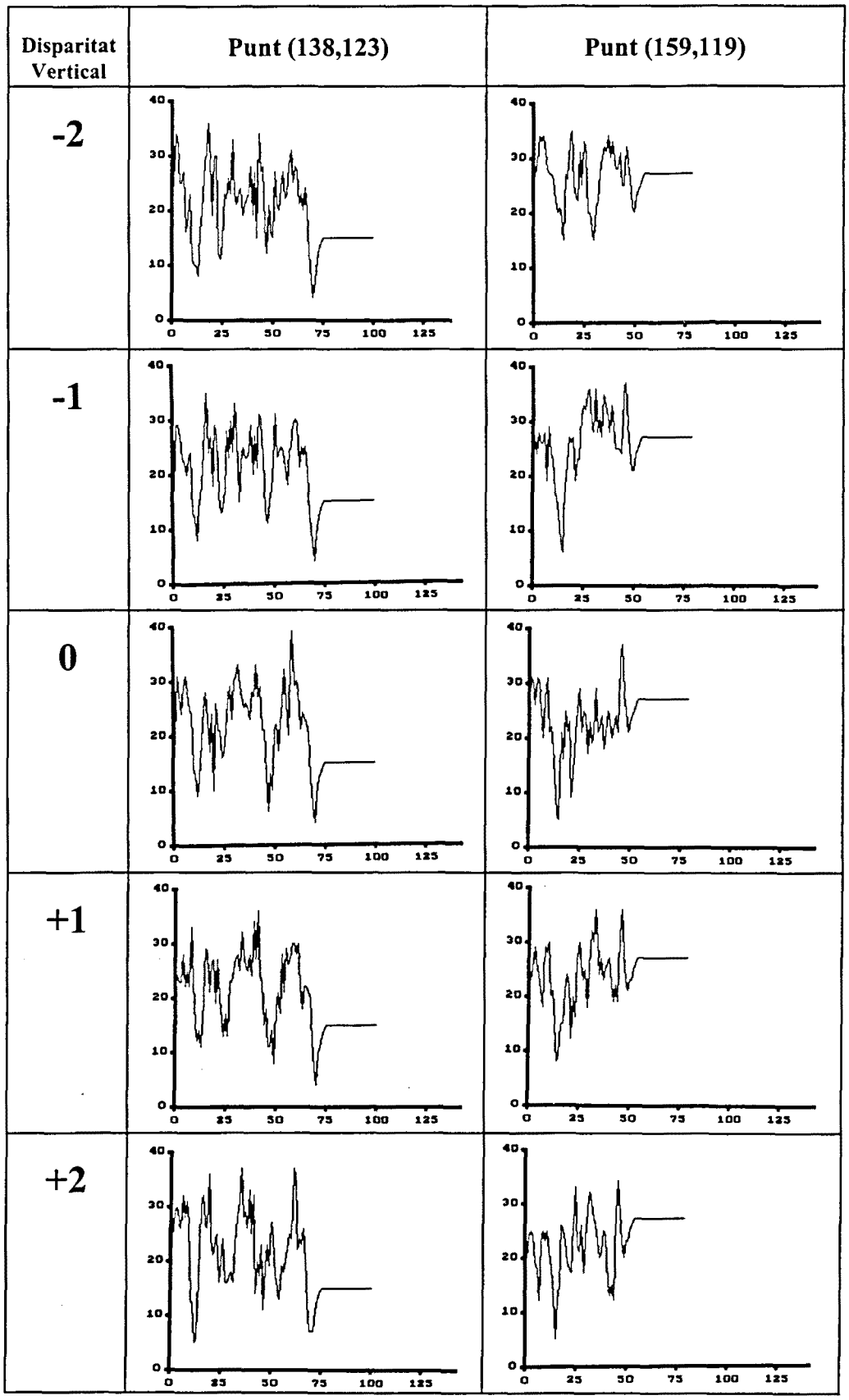


Figura 3.39. (b). Evolució de la funció D(j) per diferents linies d'imatge al voltant de la linia epipolar. Es pot observar el manteniment del punt on D(j) es fa mínima.

3.4.5. Inicialització del mòdul de seguiment.

La inicialització del sistema consisteix en elegir i determinar l'àrea d'interès sobre la que es vol realitzar el seguiment. Una vegada definida aquesta finestra és efectuada la selecció i aparellament automàtic de les característiques locals presents a la imatge esquerra, obtenint un nou conjunt de punts que pertanyen a característiques locals de la imatge dreta.

D'entre totes aquestes parelles de punts cal seleccionar el subconjunt de parelles que millor resultat han donat en el seu aparellament utilitzant els valors de la funció distància obtinguts per cadascuna d'elles.

Fet això s'obtè una quantitat relativament petita de parelles que serviran com a *dades d'entrada* del mòdul de seguiment. Concretament, la informació facilitada consta d'una llista de característiques locals per cadascuna de les imatges (esquerra i dreta). Cada característica local queda identificada per:

- una posició (x,y) dins de la imatge.
- un vector característic consistent en la transformada polar discreta de la regió centrada en el punt (x,y).

Aquesta informació és enviada a cada seguidor (*tracker*) respectivament, finalitzant el mòdul d'inicialització.

3.5. Seguiment bidimensional

El sistema de seguiment proposat està basat en un *reconeixement local* de la imatge, i quedaria classificat dins dels sistemes basats en el reconeixement que utilitzen operadors de baix nivell que poden ser implementats per *hardware.* Una millor relació dels mètodes de seguiment ha sigut ja presentada al final del capítol dos.

Aquests sistemes aborden el problema del seguiment com un problema de correspondència entre les característiques que defineixen els objectes a seguir en dues imatges consecutives.

En aquest sentit, els algorismes són molt semblants a aquells fets servir en l'aparellament estereoscòpic. El tipus de problemes que podem trobar, però, és diferent.

En el cas de l'aparellament estereoscòpic es feia ús de dues restriccions fortes donades per la geometria del propi sistema. La primera, el fet de que el punt homòleg es trobava a la mateixa línia de imatge en ambdues cameres. La segona, que no hi havia rotació respecte l'eix òptic entre les projeccions de les característiques locals de les dues cameres.

Al seguiment bidimensional, per contra, cal suposar que l'objecte es pot moure per tota la superfície de la imatge i a més a més pot rotar. Això sí, es fixa una limitació en la seva velocitat i acceleració (de translació i gir) que delimitarà l'àrea de bùsqueda dins de la imatge (anomenada *finestra de seguiment*).

Com a avantatge respecte a l'aparellament estereoscòpic, podem dir que al seguiment (sempre que la freqüència de mostreig sigui prou elevada) els canvis en la projecció de la característica local són petits i es produeixen de forma suau al llarg de la seqüència. No hem de fer front, en principi, a canvis sobtats en la forma de l'objecte com passa de vegades a l'aparellament estereoscòpic degut a la perspectiva.

3.5.1. Reconeixement local

De la mateixa forma que a l'aparellament estereoscòpic, per realitzar el reconeixement local s'utilitzen com a atributs a comparar, els valors dels radis de la transformació polar del contorn descrita en aquesta tesi. Els valors dels radis actuen, per tant, com a vector característic (o *token)* del sistema de reconeixement.

El reconeixement local s'efectua avaluant una funció similitud S definida com:

$$S(i,j) = MAX\_D - D(i,j) \qquad [\text{Eq. } 3.7]$$

on D és una funció distància (o error) i MAX_D és el valor màxim que pot agafar aquesta funció.

La funció distància D triada ha sigut la mateixa que es fa servir en la fase d'inicialització (Eq. 3.6):

$$D(i,j) = \sum_{\theta=0}^{7} \left[r(\theta) - m_{ij}(\theta) \right]^2 \quad \forall (i,j) \in \textit{finestra de seguiment} \qquad \text{[Eq. 3.8]}$$

D(i,j) és una funció discreta definida dins de la finestra de seguiment, o sigui, al voltant de la posició on s'espera trobar la característica local seguida. Els valors de la funció *D(i,j)* representen l'error entre el vector característic de la regió de la imatge al voltant de la característica local seguida $r(\theta)$, que actua com a patró, i els vectors característics associats a cada pixel dins de la finestra de seguiment $m_{ij}(\theta)$.

La posició *(i,j)* on la funció de similitud *S(i,j)* arriba al seu valor màxim és agafada com la nova posició del patró seguit.

La finestra de seguiment a part de limitar la bùsqueda i per tant baixar el cost de l'aparellament, serveix també com a filtre quan hi han diferents objectius a seguir dins de la mateixa imatge. La dimensió d'aquesta finestra ha de ser llavors el més petita possible, però no sempre és vàlida una finestra massa petita. Com va ser exposat al capítol dos, la dimensió de la finestra ve fixada bàsicament per la velocitat (i acceleració) relativa que presenti la projecció de l'objectiu dins de la imatge que és funció de la freqüència de mostreig del sistema de seguiment.

A fi d'avaluar la fiabilitat d'aquesta funció d'error *D(i,j)* s'han realitzat l'anàlisi dels següents resultats:

1. Comparació entre les transformacions polars discretes pertanyents a diferents figures de contorn tancat i de format similar. Això ens permetrà avaluar el poder de discriminació de la funció entre diferents formes que es puguin trobar a la imatge.

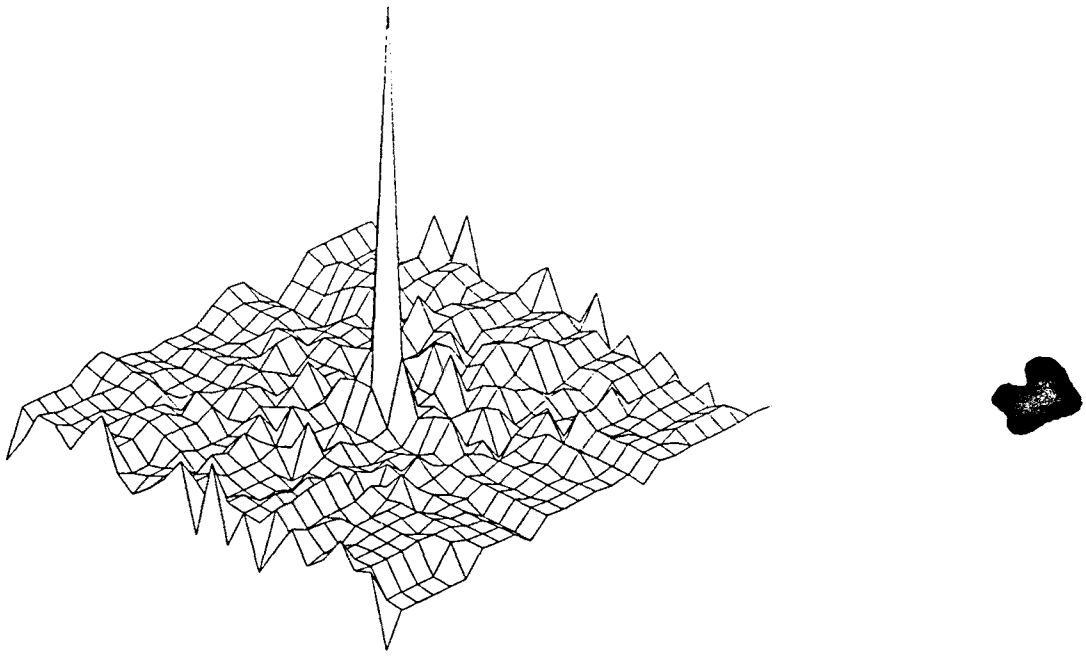
   Els resultats es mostren a l'annex 2.

2. Evolució de la funció similitud (*S*) en la cerca d'una certa figura dins d'una imatge (x,y) en presència d'altres figures de format similar.

   Els resultats obtinguts es mostren a les figures 3.40 i 3.41.

3. Evolució de la funció similitud (*S*) en front la rotació. S'ha analitzat la cerca d'un vertex de 90° sense girar (x,y) i girat esquerra i dreta fins a 180° (x,α) i (y,α). Es mostra també l'evolució de la funció *S* utilitzant una transformació de 16 radis (normalitzada). Com es pot apreciar a la figura 3.42, l'increment en el nombre de radis no afegeix una millora considerable, com ha sigut demostrat de forma teòrica a l'apartat 3.2.3.

   Els resultats obtinguts es mostren a la figura 3.42.



(a) Funció de similitud S(i,j) *(b) Patró*

Figura 3.40. Detecció i localització d'un determinat patró en la imatge.

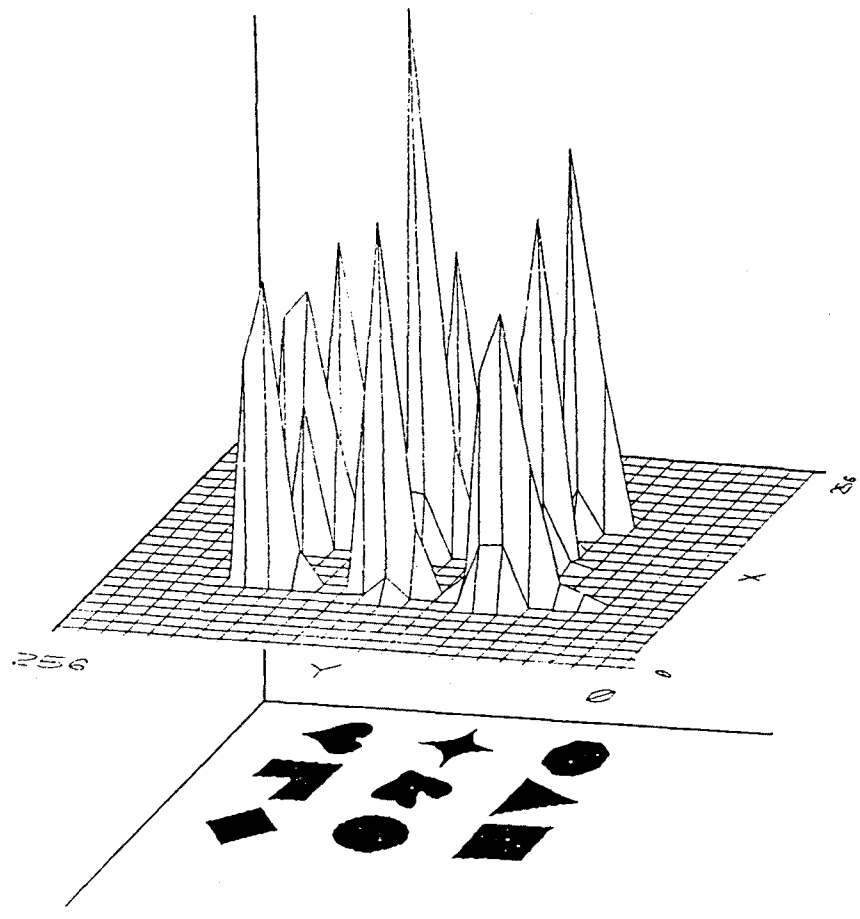
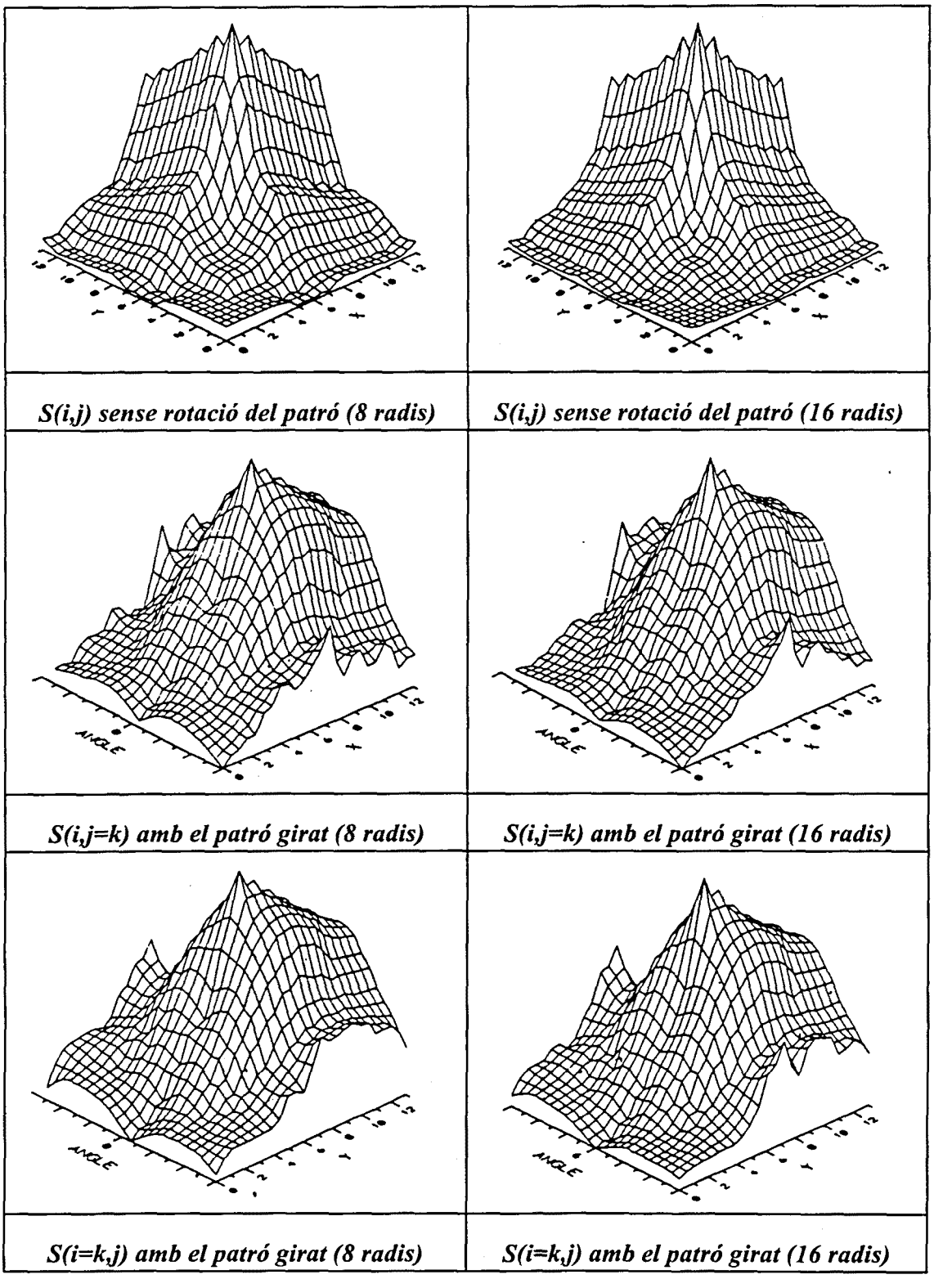


Figura 3.41. Funció de similitud S(i,j) entre un patró arbitrari (posició central) i altres vuit patrons de mida similar presents en la mateixa imatge.



S(i,j) sense rotació del patró (8 radis) | *S(i,j) sense rotació del patró (16 radis)*

S(i,j=k) amb el patró girat (8 radis) | *S(i,j=k) amb el patró girat (16 radis)*

S(i=k,j) amb el patró girat (8 radis) | *S(i=k,j) amb el patró girat (16 radis)*

Figura 3.42. Evolució de la funció similitud S(i,j) en front la rotació d'un patró. La figura (a) mostra el resultat amb el patró sense girar. Les figures (b) i (c) mostren el resultat amb el patró girat fins a ±180º. A la columna de la dreta es mostren els mateixos resultats aplicant una transformació polar de 16 radis.

3.5.2. Solució al problema de la rotació.

A l'apartat anterior s'ha fet una descripció de la solució donada al problema del seguiment basada en l'establiment d'una funció distància *D(i,j)* entre la forma esperada i l'obtinguda a cada punt *(i,j)*, que tal i com ha sigut definida només tenia en compte la possible translació de la característica local buscada dins de la finestra de seguiment. El fet és, però, que aquesta característica local pot girar també respecte a l'eix òptic de la camera durant el seu seguiment.

Aquest problema de la rotació és un dels més adversos dins del seguiment automàtic d'objectius mitjançant visió per computador [Kanatani, 94]. És en aquest punt on la descripció polar de la imatge presentada en aquesta tesi, juga un paper fonamental en l'estalvi de temps necessari per efectuar l'aparellament entre punts homòlegs. La transformada polar del contorn utilitzada, converteix els girs a l'espai X,Y en una translació a l'espai ρ,θ.

Degut a la discretització que s'ha fet d'aquesta transformació polar dels contorns (vuit radis), aquesta representació no és sensible a rotacions de l'objectiu massa petites tal i com es mostra a les figures 3.43 i 3.44.

Els resultats obtinguts en la evolució de l'error d'aparellament en funció de l'angle de rotació per vèrtexs de diferents angles (30°, 60°, 90° i 120°), s'expressen en termes relatius (%) respecte a l'error màxim d'aparellament *MAX_D* que es possible obtenir.

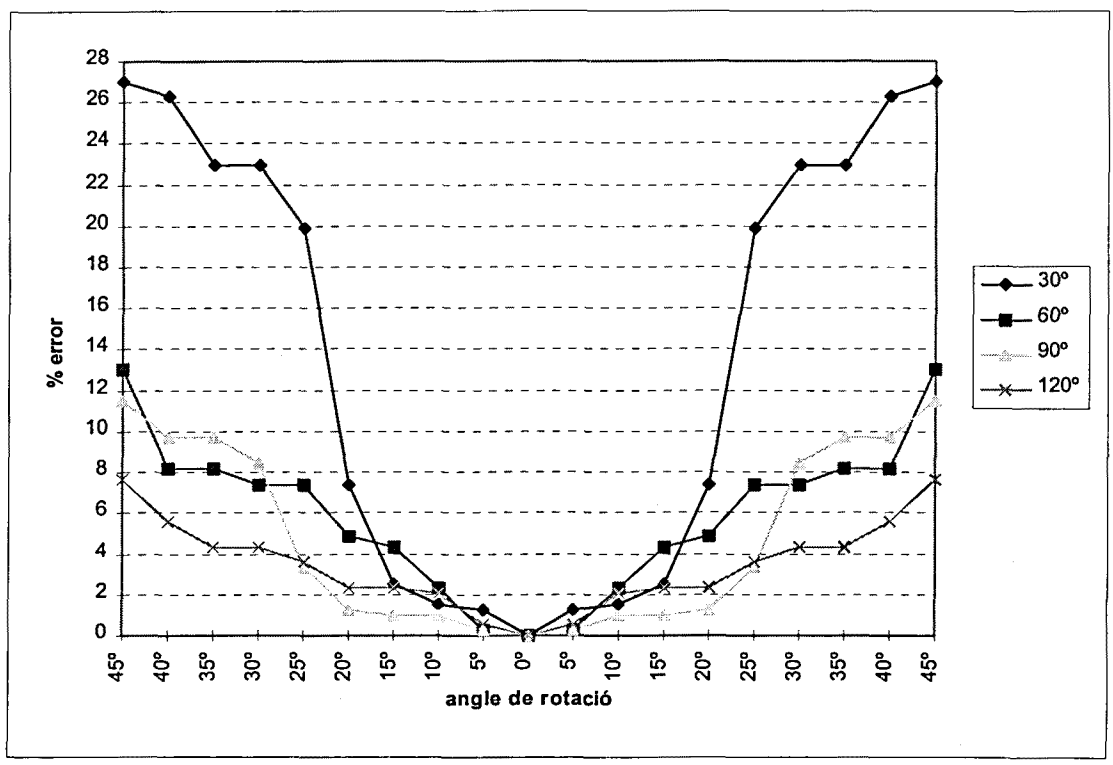


Figura 3.43. Evolució de l'error d'aparellament en funció de la rotació per vèrtexs de diferents angles.

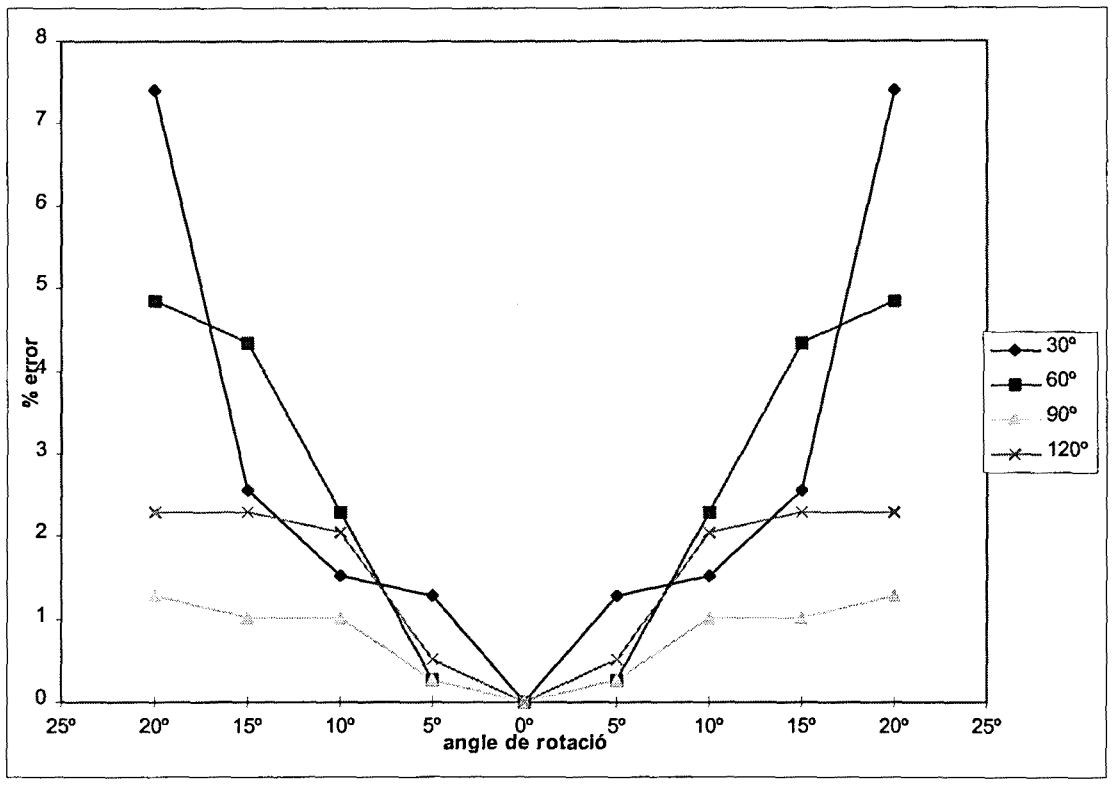


Figura 3.44. Detall de la zona de rotacions < 20º

A continuació es descriuen dos dels mètodes provats per fer front a aquest problema de l'associació de dades davant de les rotacions de les característiques locals durant el seguiment.

3.5.2.1. Normalització angular del vector descriptor.

Per garantir el seguiment dels objectius quan aquests giren, ha sigut provada una normalització angular clàssica de la representació polar d'un contorn davant dels girs [Gonzalez,87][Frau,91]. Aquests mètodes proposen que el radi amb valor més gran marqui l'inici de la representació, o sigui s'assigni al radi 0. En el cas que hi hagi més d'un radi amb aquest mateix valor s'agafa el primer de la sèrie. D'aquesta forma obtenim una invariància a la rotació a l'hora d'aplicar la funció distància.

La funció distància utilitzada queda:

$$D(i,j) = \sum_{\theta=0}^{7} \left[r(\theta) - m_{ij}(\theta + offset_{ij}) \right]^2 \quad \forall (i,j) \in finestra\ de\ seguiment$$

On $r(\theta)$ és la transformació polar del patró ja normalitzada.

Evidentment, $m_{ij}(\theta)$ ha de ser normalitzada seguint el mateix criteri a cada nou pixel (i,j) per poder efectuar la comparació. El valor de $offset_{ij}$ ve donat per la posició relativa al radi 0, que ocupa el radi amb valor més gran dins de la transformació polar associada al pixel *(i,j)* ($m_{ij}(\theta)$). La suma de θ i $offset_{ij}$ és una suma circular en mòdul 8.

Les proves realitzades aplicant aquest mètode al seguiment d'objectius que giren no han sigut totalment satisfactòries. La conclusió que poden treure és que la utilització d'aquest mètode clàssic de normalització de la transformada polar del contorn davant de la rotació, funciona en casos on la descripció dels contorns sigui tancada i amb prou mostres radials com per que una rotació no afecti de forma greu la pròpia descripció.

En el nostre cas la descripció del contorn es veu molt alterada davant de rotacions degut a dues raons:

- la baixa resolució que té la transformació polar proposada, tant en el nombre de radis com en el ventall limitat de valors que aquests poden agafar.

- el fet de voler seguir característiques locals amb contorns oberts (vol dir radis sense informació de distància associada) que agrava l'anterior circunstància.

Aquests fets dificulten que la normalització descrita anteriorment presenti estabilitat davant de la rotació. Petites rotacions de la característica local seguida pot provocar canvis en la seva descripció, que afegits als canvis deguts al soroll en la localització dels pixels de contorn, poden resultar prou importants com per donar un resultat erroni en la normalització i per conseqüent un resultat erroni en el càlcul de la funció distància utilitzada per efectuar el seguiment.

A més a més, aquesta normalització angular de la transformació polar respecte als girs té associada una pèrdua de discriminació entre característiques semblants que pertanyin a objectes diferents i que apareguin a prop en la imatge, tot i amb un angle de rotació molt diferent.

Això i l'alt cost computacional que suposa la necessitat de normalitzar angularment totes i cadasquna de les transformacions polars donades pel mòdul de processat d'imatge ($m_{ij}(\theta)$), fa que s'hagi buscat un altre mètode alternatiu a aquesta normalització clàssica en la implementació final del sistema de seguiment.

3.5.2.2. Comparació múltiple.

El mètode finalment implementat al sistema de seguiment per evitar els problemes causats per la rotació del patró dins de la seqüència d'imatges, consisteix en la comparació simultània de les transformacions polars dels pixels de la finestra de seguiment amb el patró buscat i el patró girat a esquerra i dreta 45°. Resulta evident que no cal realitzar el gir esquerra i dreta del patró en 45°, ja que coincideix amb un canvi d'índex en l'accés a la informació de la seva transformada polar. Això

representa un avantatge clar de l'us de la transformada polar respecte d'altres mètodes de descripció de contorns o de seguiment basat en la correlació directa de imatges (*"pattern maching"*) com és el cas del presentat a [Inoue,93]

La distància mínima que surti de la comparació, per cada pixel dins de la finestra de seguiment, de les transformacions polars dels contorns de la imatge amb aquests tres patrons determinarà la nova posició (i,j) del objectiu i també la seva orientació actual (discretitzada en 45°) [Amat,93a].

El procediment és com segueix:

offset = 0; /* *al moment de la inicialització* */
per *cada nova imatge* ***fer***
Dmin= distància màxima
per *tot (i,j)* $\in$ *finestra de seguiment* ***fer***
calcular $m_{ij}(\theta)$
avaluar les tres distàncies:

$$D_{-1}(i,j) = \sum_{\theta=0}^{7} (r(\theta + offset - 1) - m_{ij}(\theta))^2$$

$$D_{0}(i,j) = \sum_{\theta=0}^{7} (r(\theta + offset) - m_{ij}(\theta))^2$$

$$D_{+1}(i,j) = \sum_{\theta=0}^{7} (r(\theta + offset + 1) - m_{ij}(\theta))^2$$

D = min (D_{-1}, D_0, D_{+1})
Si *D < Dmin* ***llavors***
Dmin=D
nova (x, y) = (i , j)
Si *D= D*$_{-1}$ ***llavors*** *nou _offset=offset-1*
Si *D= D*$_{+1}$ ***llavors*** *nou_offset=offset+1*
fper /* *finestra de seguiment* */
offset = nou_offset
retorna (Dmin, nova (x,y), nou_offset)
fper /* *nova imatge* */

O sigui, que l'error màxim degut a la rotació serà de ± 22,5°. (Veure la figura 3.46). Passat aquest llindar de rotació el valor d'*offset* serà actualitzat (correcció de 45° en l'accés a la informació del patró a comparar). L'actualització del valor d'*offset* és sempre en mòdul 8.

Un clar avantatge respecte al mètode proposat a l'apartat anterior és el fet de no haver de normalitzar les mostres adquirides dins de la finestra de seguiment. Per contra s'han d'avaluar tres funcions distància amb el mateix patró (+offset). Aquestes comparacions han estat optimizades de forma que es fa el seu càlcul a través de taules magatzemades en memòria (*"look-up tables"*). De fet és fins i tot factible la seva paral.lelització en una implementació *hardware* de baix cost. (La descripció de la implementació *hardware* es tractada al capìtol 4 d'aquesta tesi).

Com a exemple dels resultats de l'aplicació de les tres comparacions sobre una mateixa imatge, es presenta una seqüència amb un gir d'un vèrtex de 90° en una determinada direcció.

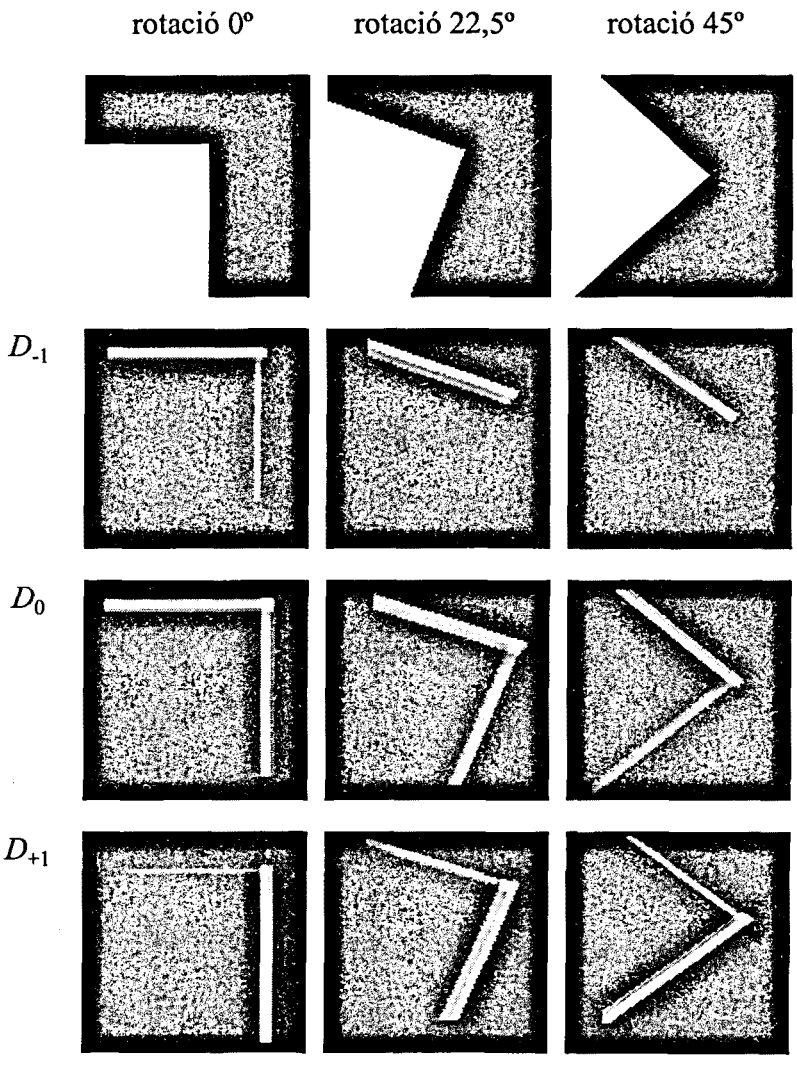


Figura 3.45. Resultat de la comparació amb el patró (D_0) i el patró girat a l'esquerra -45° (D_{-1}) i a la dreta +45° (D_{+1}), al llarg d'una seqüència.

Les gràfiques de la figura 3.46 mostren l'evolució de l'error d'aparellament en funció de la rotació, per a diferents vèrtexs, tenint en compte la normalització proposada. Els valors de l'error (coordenada Y) s'expressen en termes relatius (%) respecte a l'error màxim d'aparellament *MAX_D* que es possible obtenir.

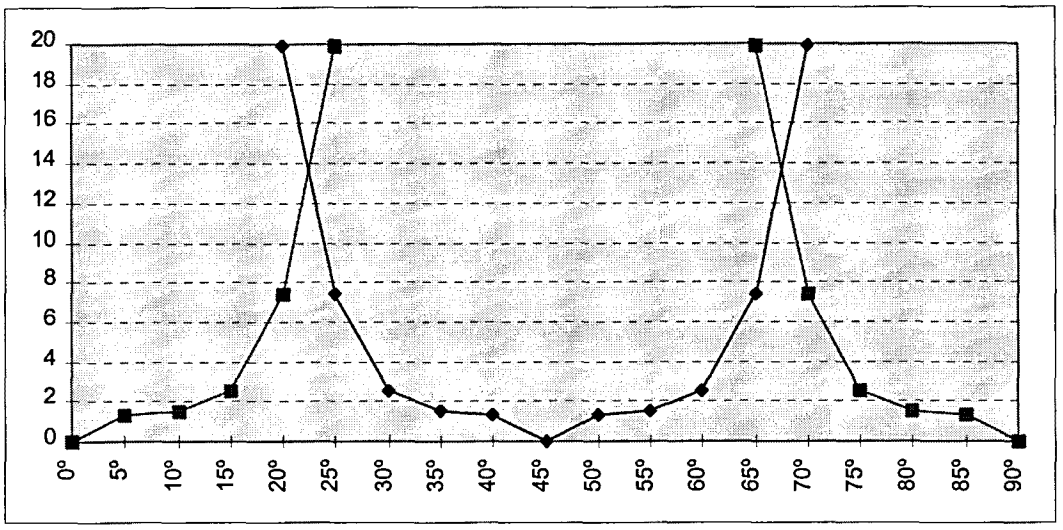


Figura 3.46 (a). Evolució de l'error d'aparellament al llar d'un gir de 90º per a un vèrtex de 30º

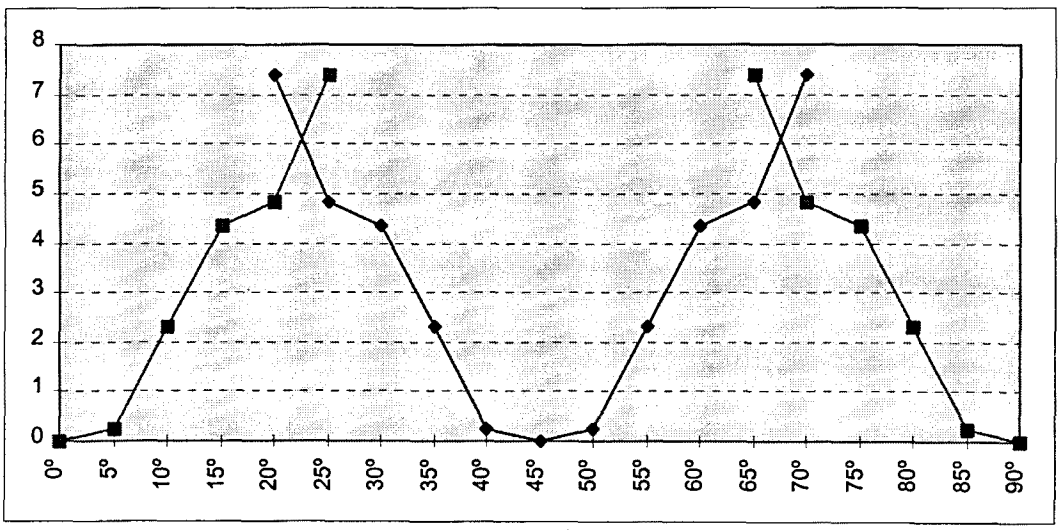


Figura 3.46 (b). Evolució de l'error d'aparellament al llar d'un gir de 90º per a un vèrtex de 60º

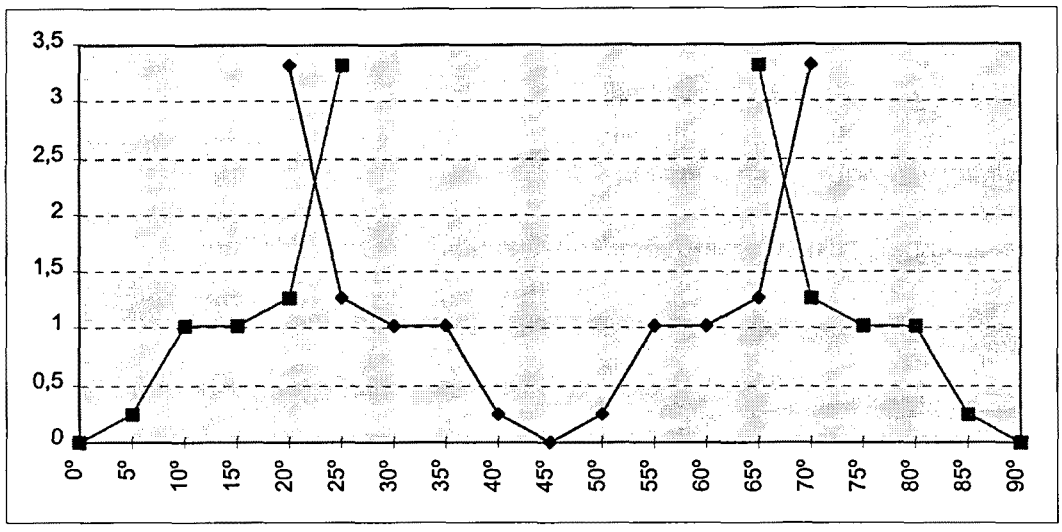


Figura 3.46 (c). Evolució de l'error d'aparellament al llar d'un gir de 90° per a un vèrtex de 90°

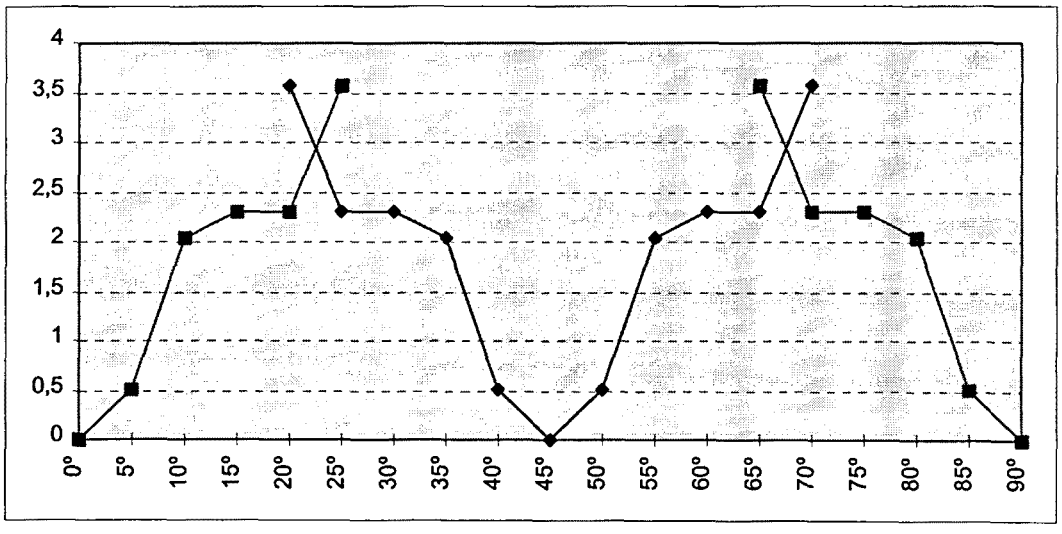


Figura 3.46 (d). Evolució de l'error d'aparellament al llar d'un gir de 90° per a un vèrtex de 120°

Com es pot comprovar a partir dels resultats, amb la normalització proposada l'error d'aparellament en cap cas supera un valor prou significatiu com per arribar a perdre l'objectiu seguit, o confondre'l amb un altre objectiu de diferent transformació polar.

El pitjor cas es dona amb els vertexs molt tancats (30°) on s'obté un percentatge d'error igual al 14% per un gir de 22,5°. Superat aquest angle de gir la normalització proposada fa baixar l'error d'aparellament fins anular-ho als 45°.

3.5.3. Aplicacions del seguiment bidimensional.

El seguiment bidimensional, representa un cert interès tecnològic, donades les elevades possibilitats d'aplicació de tot tipus que pot donar lloc. Es presenta a continuació alguns dels múltiples assaigs que s'han realitzat a fi de provar i avaluar la fiabilitat del sistema de seguiment bidimensional proposat i que ja han sigut presentats en diferents congressos internacionals. S'inclouen dos aplicacions prou diferents, una en un entorn industrial i una altra en el sector serveis.

3.5.3.1 Pintat de motors en una cel.la robotitzada

Un dels primers exits en l'aplicació del sistema de seguiment consisteix en el seguiment en temps real d'uns motors que han de ser pintats per un robot en una cel·la de pintat automatitzada [Amat,92].

Els motors són transportats penjant del seu eix, de forma que a part de moure's poden girar sobre si mateixos. La camera s'ha situat sobre el transportador aeri, oferint una projecció zenital de l'escena.

Inicialment l'usuari indica al sistema la zona on apareixen els motors dins de l'escena. Per a cada nou motor, el sistema localitza de forma automàtica les característiques locals més prominents del contorn del motor (utilitzant la funció distància de l'eq. 3.4). D'entre aquestes, selecciona aquelles que són més pròximes a l'extrem superior i inferior de la imatge (per tenir millor resolució en la mesura de l'angle de gir del motor). Tot seguit comença el seu seguiment.

Les figures 3.47 a 3.50 il·lustren la disposició i funcionament del sistema:

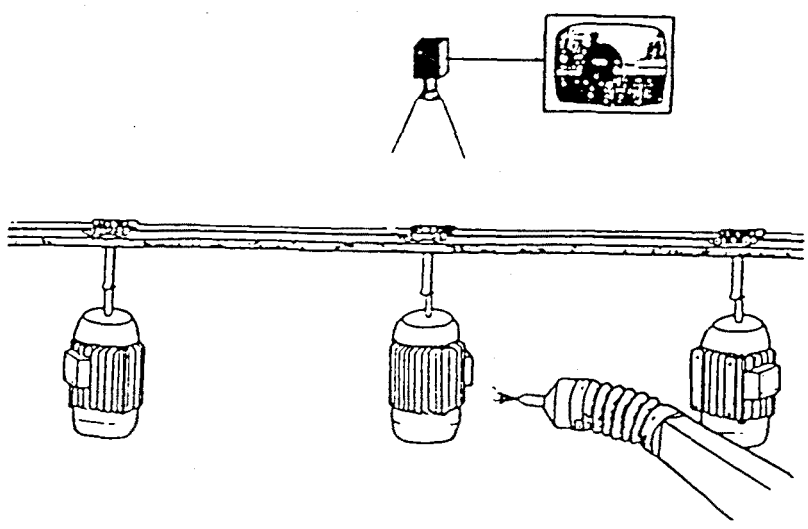


Figura 3.47. Disposició dels elements del sistema de pintat automatitzat

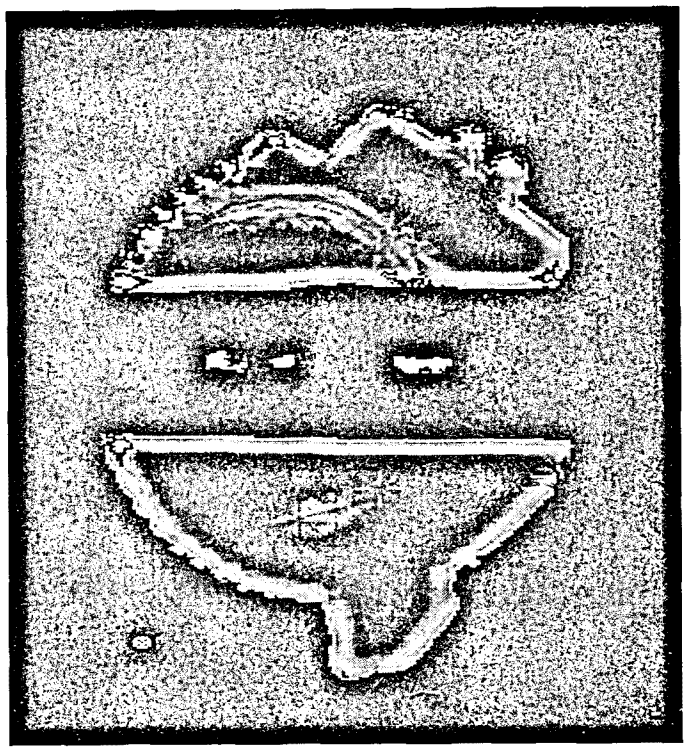


Figura 3.48. Resultat d'aplicar la funció fiabilitat descrita a l'apartat 3.3 (Eq.3.4) per a la selecció de característiques locals, dins de la zona d'imatge indicada per l'usuari.

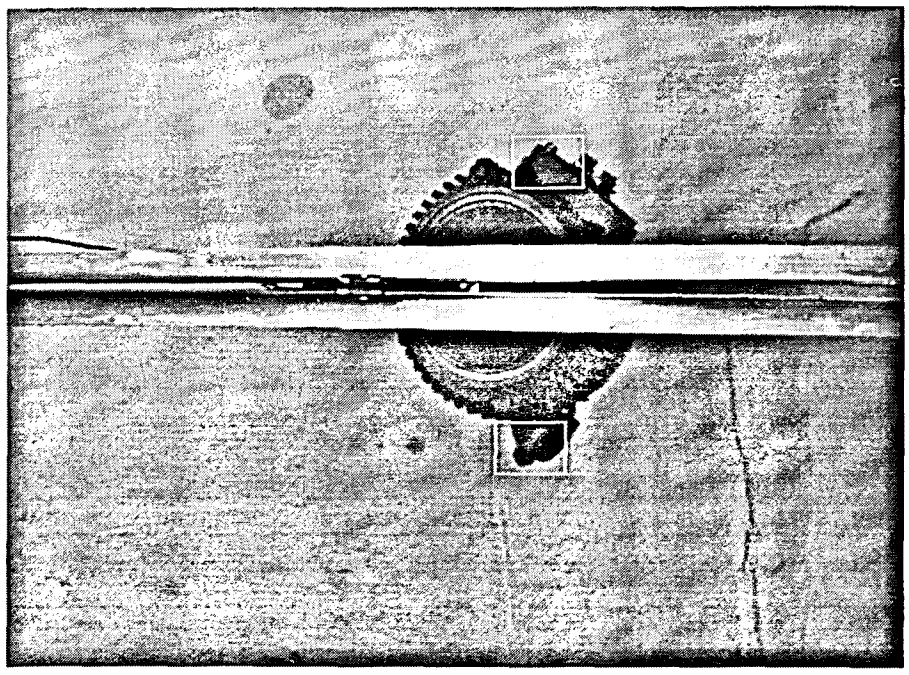


Figura 3.49. Seguiment de les característiques locals seleccionades.

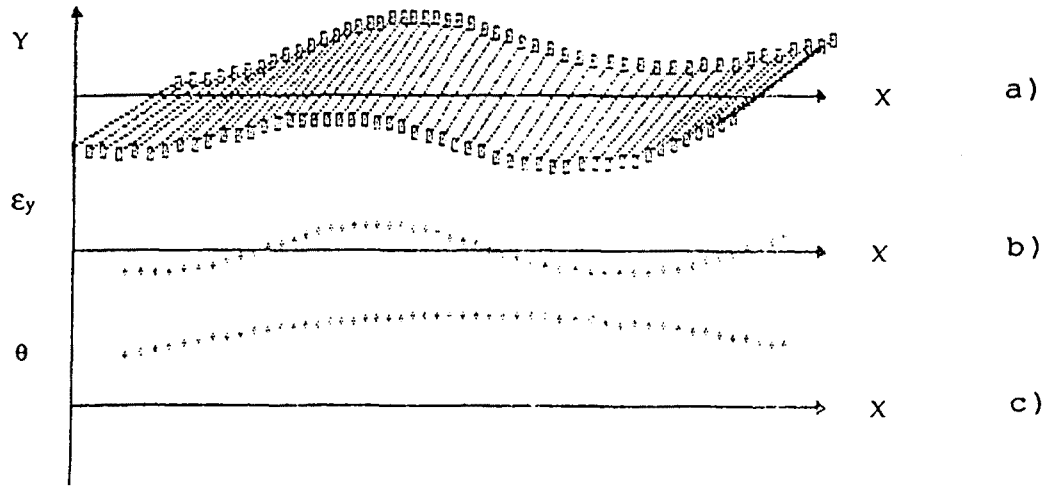


Figura 3.50. Trajectòria seguida pel motor. a) posició de les característiques locals al llarg del temps. b) desviació (oscil·lació) respecte al sistema de transport. c) Evolució de l'orientació del motor.

3.5.3.2 Seguiment de vehicles

Una altra aplicació d'interès és determinar el flux de vehicles en una cruïlla, plaça o rotonda on els vehicles tinguin opció a entrar i sortir per accessos diferents. El simple control del nombre de vehicles que passen pels accessos d'entrada i les possibles sortides no proporciona informació respecte l'origen i destí dels vehicles. Per obtenir aquesta informació cal conèixer la trajectòria individual de cada vehicle [Aranda,93].

La seqüència d'imatges de la figura 3.51 mostra l'evolució d'un seguiment. A la figura 3.52. es mostren els resultats estadístics de les trajectòries seguides pels vehicles observats.

Altres aplicacions del sistema de seguiment bidimensional proposat poden ser trobades a [Amat,93a][Aranda,94][Amat,97].

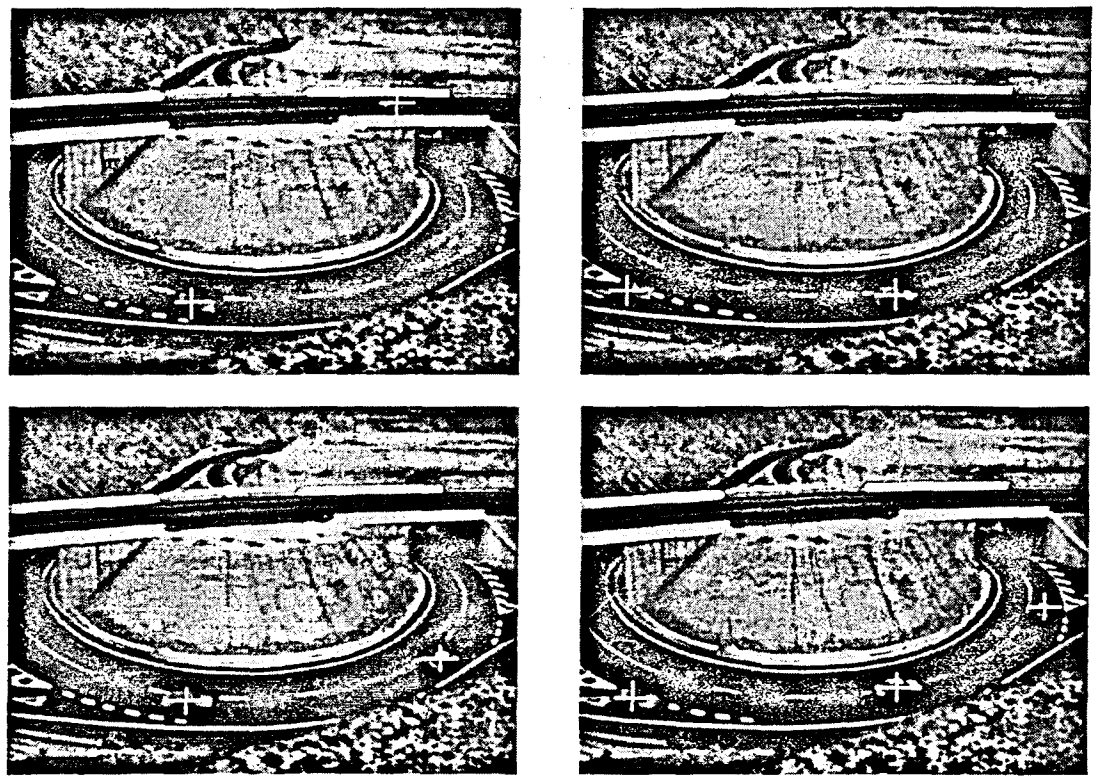


Figura 3.51. Quatre quadres d'una seqüència de seguiment de vehicles en una rotonda.

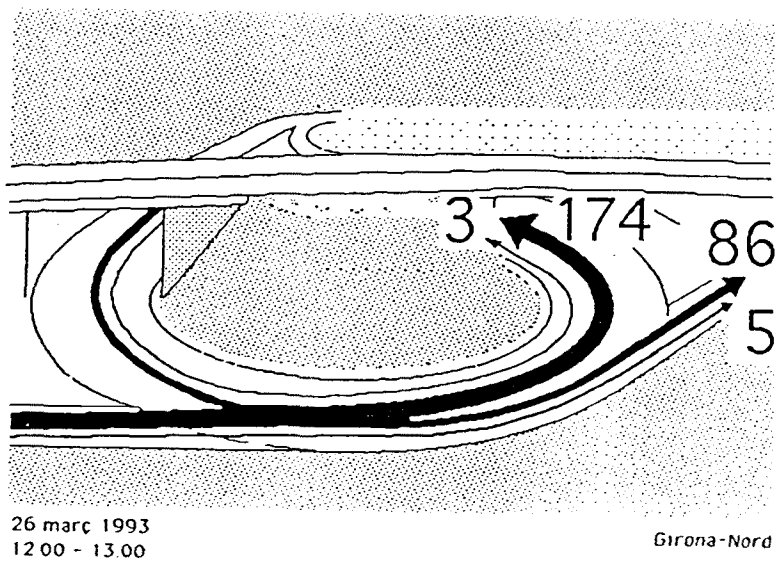


Figura 3.52. Histograma de les trajectòries seguides pels vehicles en la rotonda analitzada.

3.6. Seguiment tridimensional

Els sistemes de seguiment tridimensional d'objectius mitjançant l'estereovisió acostumen a efectuar una certa optimització en el temps de computació de l'aparellament estereoscòpic (en alguns casos disposen d'una arquitectura paral·lela per la seva implementació). En la majoria de casos, primer realitzen un aparellament binocular entre imatge esquerra i dreta, i més tard, un aparellament temporal dins de la seqüència amb les dades de posició tridimensional [Hong, 93] [Hollinghurst, 94] [Zhang, 94] [Shieh, 94].

Al sistema proposat, el seguiment bidimensional es realitza en paral·lel a les imatges esquerra i dreta sobre cadascuna de les característiques locals seleccionades. L'aparellament estereoscòpic no és llavors necessari durant el seguiment, ja que s'ha fet inicialment en l'etapa de selecció de característiques. Amb aquesta metodologia es millora la robusteça del seguiment tridimensional, que es beneficia de la millor qualitat de l'aparellament de seguiment respecte a l'aparellament estereoscòpic. Aquesta característica ja va ser exposada a l'apartat 3.5 d'aquest capítol. Evidentment, tal i com fan alguns autors [Ho, 96], sempre és possible fer a més a més l'aparellament estereoscòpic per obtenir informació redundant que avaluï la fiabilitat del seguiment.

El seguiment bidimensional realitzat sobre les imatges esquerra i dreta, proporcionen la següent informació relativa a l'estat de cada característica local a cada nova parella de imatges :

1. La posició, *Pe= (xe , ye)* i *Pd= (xd , yd)* dins de la imatge esquerra i dreta respectivament , on ha sigut localitzada la característica local.

2. La fiabilitat, *Se i Sd* , d'aquesta localització, obtinguda a partir de la funció de similitud *S* utilitzada per fer l'aparellament temporal.

Siguin ***Te*** i ***Td*** les trajectòries esquerra i dreta de una determinada característica local de la imatge al llarg d'una seqüència de *n* imatges, podem representar-les com:

$$Te_t = (Pe_1, Pe_2, ..., Pe_n)$$

$$Td_t = (Pd_1, Pd_2, ..., Pd_n)$$

A partir d'aquestes dades, i utilitzant la geometria coneguda del sistema estereoscòpic, s'obté la trajectòria de cadascuna de les característiques seguides a l'espai tridimensional.

$$T_t = (X_1, X_2, ..., X_n)$$

on cada X_i és la posició *(x,y,z)* , a cada nova imatge, de la característica local seguida dins del sistema de coordenades absolut del nostre sistema estereoscòpic.

3.6.1. Obtenció de la posició a l'espai

Per cadascuna de les característiques locals seguides, i coneguda la posició de les seves projeccions dins de les imatges esquerra i dreta, *Pe=(xe,ye) i Pd=(xd,yd)* respectivament, és possible calcular la posició tridimensional de la característica local respecte a un sistema de coordenades centrat entre les dues cameres. La geometria del sistema ha estat descrita a l'apartat 3.4 d'aquest capítol.

Seguint el model de camera *pin-hole,* el sistema estereoscòpic satisfà la següent igualtat entre triangles (Eq. 3.5):

$$z = \frac{b.f}{(xe - xd)}$$

Una vegada coneguda z es verifiquen les següents identitats:

$$x = \frac{xe.z}{f} - \frac{b}{2} \qquad\qquad y = \frac{ye.z}{f}$$

La coordenada x té un desplaçament de b/2 donat que s'ha considerat l'origen del sistema al mig de les dues cameres (Figura 3.32).

Les coordenades x i y, poden ser expressades directament en funció de la disparitat, que ha estat obtinguda a partir de l'aparellament, evitant el càlcul indirecte a través de la z i per tant els error deguts al seu càlcul:

$$x = \frac{xe.b}{(xe - xd)} - \frac{b}{2} \qquad y = \frac{ye.b}{(xe - xd)} \qquad z = \frac{b.f}{(xe - xd)} \qquad \text{[Eqs. 3.9]}$$

Aplicant les equacions anteriors obtenim a cada nova imatge la posició $X = (x,y,z)$ de cada característica local.

3.6.2. Anàlisi de la coherència de les dades

En presència de múltiples objectius amb moviment independent, el mètodes de seguiment han d'incorporar una sèrie de restricions en les dades proporcionades, que tingui en compte les característiques dels objectius i el seu moviment, per superar els possibles errors d'aparellament i assegurar la coherència en les dades referents a la trajectòria.

A diferència d'altres sistemes publicats [Bar-Shalom, 95], el sistema de seguiment proposat no requereix d'un coneixement *a priori* de la cinemàtica i dinàmica dels objectes a seguir, donat que es realitza un reconeixement continu. Es busca d'aquesta manera generalitzar la inicialització del seguiment. Es fa servir una freqüència de mostreig prou elevada com per poder garantir la presumpció de suavitat en el moviment dels objectius (*smoothness*) dins de la seqüència d'imatges.

S'espera així que durant el seguiment es trobaran els objectius en una posició propera a la que ocupaven a la imatge anterior. A més a més, degut a la inèrcia, la velocitat d'una entitat física no pot variar de forma instantània, ni en mòdul ni en direcció. Aquestes afirmacions, cal recordar, només són correctes si es compta amb una freqüència de mostreig prou elevada, com és el cas del sistema presentat, gràcies a la implementació *hardware* de la part de processat de dades.

Per tant, són tres els paràmetres que determinen la coherència:

1. L'increment de posició, $\Delta X = \Delta(x,y,z)$, entre dues imatges consecutives.

2. L'increment de velocitat entre dues imatges consecutives.

3. Els canvis en la direcció del moviment entre dues imatges consecutives.

L'anàlisi en la coherència de les dades proporcionades pel seguiment bidimensional (seguint la presumpció de suavitat) té dues finalitats. La primera és de caràcter restrictiu (filtre), imposant unes condicions al moviment i al possible canvi de forma dels objectius, que s'hauran de complir sempre. La segona, és avaluar aquest moviment donant un valor normalitzat entre 0 i 1 per indicar la seva coherència com una dada més que proporciona el sistema.

3.6.2.1. Criteris d'aceptació de les dades

Amb la finalitat de millorar el temps de càlcul, i descartar d'una forma més ràpida aquelles posicions proporcionades pel seguiment bidimensional que resulten incoherents, el sistema imposa una sèrie de restriccions al moviment dels objectius i al canvi de forma que poden presentar entre imatges consecutives. En el cas que les condicions es satisfacin, el resultat de l'aparellament és acceptat, altrament el resultat és rebutjat fins una altra mostra.

Aquestes restriccions tenen en compte:

1. les finestres de seguiment (posició i dimensió)

2. el canvi de disparitat

3. les funcions de similitud

Les finestres de seguiment marquen els límits dins de cada imatge on la característica local pot ser trobada. Amb la determinació de la seva posició i la seva mida es pot restringir la zona de búsqueda, de forma, que en cas de trobar la característica local, aquesta compleixi les condicions mínimes de coherència en la trajectòria que han sigut fixades: increment de posició, velocitat i direcció de moviment.

Ja que aquests canvis han de ser petits en la trajectòria tridimensional de l'objecte, també ho seran en les seves projeccions a les imatges esquerra i dreta. Quan més petita sigui la dimensió de les finestres de seguiment, més ràpid serà el sistema de seguiment

i millor la fiabilitat de l'aparellament. Aquest aspecte serà estudiat més detingudament a l'apartat 3.6.3.

Es pot produir, però, que un petit desplaçament x,y al pla de la imatge es correspongui amb un desplaçament gran en les z. Per això, tot i que les finestres restringeixen aquesta possibilitat, és necessari també una condició de coherència en la component z, calculada a partir de la disparitat present entre les dues projeccions sobre les cameres. La condició es basa en limitar l'increment que es produeix en la coordenada z entre dues parelles d'imatge consecutives, així com l'increment de la seva derivada (velocitat). Siguin Z_{k-1} , Z_k , Z_{k+1} les profunditats successives d'una característica local dins de la seva trajectòria, estimades a partir de la disparitat present a les parelles d'imatge *k-1*, *k* i *k+1* respectivament. Llavors definim les condicions sobre l'increment en z i la seva derivada en el moment *k+1* com

$$\Delta Z_{k+1} = |Z_{k+1} - Z_k| < \Delta Z\_max$$

$$\Delta Z'_{k+1} = |\Delta Z_{k+1} - \Delta Z_k| < \Delta Z'\_max \qquad amb\ \Delta Z'\_max = \Delta Z\_max$$

La última restricció imposada a les dades del seguiment bidimensional, considera el valor de les funcions de similitud calculades per fer l'aparellament en les imatges esquerra i dreta (*Se i Sd*) (Eq. 3.7). Aquestes funcions ens donen un valor per poder estimar la fiabilitat amb la que es produeix el seguiment. Si una d'elles no dona un bon resultat la nova posició no és acceptada com a vàlida fins al següent aparellament considerat vàlid. Cal que les dues funcions de similitud siguin prou bones, per tant, un bon estimador de la fiabilitat del nou aparellament és:

$$S = Se.Sd > Smin$$

Aquestes tres restricions (finestra de seguiment, disparitat i similitud) aplicats sobre les dades procedents del seguiment bidimensional proporcionen un mecanisme ràpid d'assegurar la coherència mínima de les dades de la trajectòria desestimant aquelles mostres que no compleixen les condicions anteriors.

3.6.2.2. Mesura de la coherència de la trajectòria

S'ha utilitzat també una funció per avaluar de forma conjunta els tres paràmetres que defineixen la coherència del moviment dels objectius (increment de posició, velocitat i direcció de moviment), mitjançant una *"funció de coherència" (FC).*

Aquesta *funció de coherència* ha sigut normalitzada de forma que el seu valor estigui comprés dins de l'interval [0,1]. Un valor proper a 1 indicarà una falta de coherència en les dades relatives a la trajectòria. Pel càlcul de la funció coherència presentada farem servir la següent notació:

Siguin X_{k-1} , X_k , X_{k+1} les posicions successives d'una característica local dins de la seva trajectòria, estimades a partir de les parelles d'imatge *k-1*, *k* i *k+1* respectivament (figura 3.53).

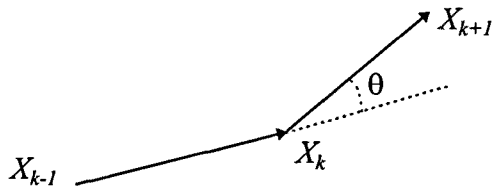


Figura 3.53. Posicions successives d'una característica local dins de la seva trajectòria

- La *coherència en la posició* de la característica local (a l'espai 3D), a l'instant *k*, ve donada en aquest cas per l'expressió:

$$CP_k = \frac{\left\|\overline{X_{k-1}X_k}\right\|}{\Delta X_{Max}} \in [0,1] \qquad \text{[Eq. 3.10]}$$

que representa el moviment relatiu de la característica entre la imatge *k-1* i la imatge *k*

- La *coherència en la velocitat,* normalitzada, ve donada per l'expressió:

$$CV_k = 2\frac{\sqrt{\left\|\overline{X_{k-1}X_k}\right\|\cdot\left\|\overline{X_kX_{k+1}}\right\|}}{\left\|\overline{X_{k-1}X_k}\right\| + \left\|\overline{X_kX_{k+1}}\right\|} \in [0,1] \qquad \text{[Eq. 3.11]}$$

que considera el quocient entre la mitja geomètrica i la mitja aritmètica del mòdul dels dos vectors desplaçament.

- La *coherència en la direcció* del moviment ve donada pel cosinus de l'angle que formen els dos vectors desplaçament (cosθ):

$$CD_k = \frac{\overline{X_{k-1}X_k}\cdot\overline{X_kX_{k+1}}}{\left\|\overline{X_{k-1}X_k}\right\|\cdot\left\|\overline{X_kX_{k+1}}\right\|} \in [0,1] \qquad \text{[Eq. 3.12]}$$

que ve donat pel quocient entre el producte escalar dels dos vectors desplaçament i el producte dels seus mòduls.

La funció coherència pot ser parametritzada ara com una funció ponderada de las anteriors funcions coherència com:

$$FC_k = w1.CP_k + w2.(1- CV_k) + w3.(1- CD_k) \quad \in [0,1] \qquad \text{[Eq. 3.13]}$$

On *w1, w2 i w3* actuen com a pesos a ponderar segons l'aplicació. La suma d'aquests pesos ha de ser igual a 1.

Aquests pesos poden ser establerts manualment segons l'aplicació on es pretengui utilitzar el sistema de seguiment, o poden ser obtinguts de forma experimental en una

fase prèvia d'experimentació en un entorn donat, observant les característiques cinemàtiques i dinàmiques dels objectes a seguir.

Al llarg de la trajectòria es pot anar obtenint un ùnic valor representatiu de la coherència a partir dels valors parcials calculats a cada nou instant k (com a mínim calen $n=3$ mostres):

$$FC_{total} = \frac{\sum_{k=2}^{n-1} FC_k}{n-2} \qquad \text{[Eq. 3.14]}$$

Aquest valor associat a cada trajectòria és donat com a dada de sortida del sistema de seguiment juntament amb les dades corresponents a les posicions $(X_1, X_2, \ldots, X_n)$. Aquest valor aporta una messura de *fiabilitat* de les dades dins de l'interval [0,1]. Quan més proper sigui aquest valor a zero, més coherents són les dades relatives a la trajectòria.

3.6.3. Control de la finestra de seguiment (predicció).

El seguiment bidimensional necessita, com a dades d'entrada, la dimensió i posició de la finestra de seguiment de cadascuna de les característiques locals seguides a fi d'efectuar l'aparellament d'aquestes. A més a més, cada característica local va acompanyada del seu vector descriptor (transformada polar). Aquesta informació es obtinguda inicialment pel mòdul d'inicialització. Durant el seguiment tridimensional, serà el mòdul d'anàlisi de dades el que determini la següent posició i dimensió de les finestres de seguiment.

3.6.3.1. Predicció de la posició de la finestra.

Sigui $T_{t=k} = (P_1, P_2, \ldots, P_k)$ la trajectòria d'un objectiu dins d'una seqüència d'imatges.

El problema de la predicció consisteix en estimar quina és la posició a la imatge $k+1$ (F_{k+1}) amb més probabilitat de que es trobi l'objectiu. Aquesta posició F_{k+1} marcarà el centre de la finestra de seguiment a la imatge $k+1$.

Si el model de moviment de l'objectiu és conegut (vol dir es coneix amb prou aproximació quina serà la P_t de l'objectiu $\forall t$) o bé és estimable a partir de les mostres precedents, n'hi ha prou amb aplicar el model per obtenir P_{t+1} com una bona aproximació. En aquest cas quedaria:

$$F_{k+1} = P_{k+1}$$

Ara bé, si l'objectiu no és conegut a priori o no presenta moviment uniforme (falta de coherència), la millor predicció és no fer-ne, i designar com a centre de la finestra de seguiment, la posició que ocupava en la imatge anterior la característica local seguida. En aquest cas quedaria:

$$F_{k+1} = P_k$$

En el nostre cas s'ha obtat per una predicció lineal de primer ordre que comtempli la posibitat d'optimitzar el seguiment d'objectes amb moviment uniforme en velocitat i direcció.

$$F_{k+1} = P_k + Q.(P_k - P_{k-1}) \quad \text{amb } Q \text{ entre 0 i 1}$$

Q pot ser donada per l'usuari del sistema, en funció de l'aplicació, o bé pot ser definida de forma automàtica a partir de la coherència que presenti la característica local al llarg de la seva trajectòria. Inicialment, i per defecte, Q es fa igual a 0.3 .

Convé observar com la predicció en la posició de les finestres tant a la imatge esquerra com a la imatge dreta es fa a partir de les dades de posició relatives a aquestes imatges i no a partir d'una predicció de la posició tridimensional i una posterior projecció d'aquesta predicció sobre la parella d'imatges, com per exemple a [Hong,93]. Aquesta metodologia representaria un cost adicional a l'hora de calcular la projecció de la posició estimada sobre cadasquna de les imatges, per la qual cosa va ser desestimada.

3.6.3.2. Predicció de la dimensió de la finestra.

Tal com ha sigut considerat anteriorment, la finestra de seguiment ha de ser el més petita possible a fi de baixar el temps de processat que el seguiment bidimensional necessita. La dimensió de la finestra de seguiment, però, ha de ser prou gran com per incloure la posició on es troba l'objectiu seguit. La finestra de seguiment marca en realitat els marges d'error esperats per la predicció del moviment. Per objectius fortament maniobrables (sense model dinàmic aparent) la dimensió de la finestra haurà de ser gran, mentre que per moviments amb velocitat uniforme podria ser molt petita (sempre que no s'esperin canvis sobtats).

Al sistema implementat s'ha optat per tenir tres dimensions de finestra fixes: 8x8, 16x16 i 32x32 pixels respectivament. Fins i tot la mida més gran (32x32 pixels), és petita comparada amb la d'altres sistemes de seguiment basat en el reconeixement. La reducció de la dimensió de la finestra de seguiment no ha estat deguda, com és habitual en altres sistemes, per un anàlisi exhaustiu de la dinàmica dels objectius i una bona predicció, si no gràcies a una implementació *hardware* que permet freqüències de mostreig superiors a les habituals en altres sistemes de seguiment basats en processadors d'imatge convencionals.

La dimensió de la finestra es selecciona en funció de l'estabilitat que presenta la posició i la velocitat aparent (dins de la imatge) de la característica local seguida, de

forma que a partir de l'error de predicció (error entre la posició de predicció (F_k) i la posició de la característica local (P_k) tenim:

Si $|P_k - F_k| \leq 2$ *pixels* *llavors dimensió_finestra = 8x8.*

Si $2 \text{ pixels} \leq |P_k - F_k| \leq 6$ *pixels* *llavors dimensió_finestra = 16x16.*

Si $|P_k - F_k| \geq 6$ *pixels* *llavors dimensió_finestra = 32x32.*

D'aquesta forma s'estableix una relació entre la predició de posició de la finestra de seguiment i la seva dimensió. Com es pot observar, s'ha permés una histèresi d'un pixel abans de fer un canvi en la dimensió.

Si recordem la fòrmula per la predicció de posició a l'instant *K:*

$$F_k = P_{k-1} + Q.(P_{k-1} - P_{k-2})$$

Per $Q=0$ (no predicció) tenim $F_k = P_{k-1}$ i llavors:

$$|P_k - F_k| = |P_k - P_{k-1}| = |velocitat_k|$$

o sigui, que la dimensió de la finestra de seguiment depèn de la velocitat aparent que presenta l'objecte sobre la imatge.

Per $Q=1$ (predicció lineal amb velocitat constant) tenim $F_k = P_{k-1} + P_{k-1} - P_{k-2}$

i llavors:

$$|P_k - F_k| = |P_k - (P_{k-1} + P_{k-1} - P_{k-2})| = |(P_k - P_{k-1}) - (P_{k-1} - P_{k-2})| =$$

$$= |velocitat_k - velocitat_{k-1}| = |acceleració_k|$$

o sigui, que la dimensió de la finestra de seguiment depèn de l'acceleració que presenta l'objecte.

3.6.3.3. Extrapolació de la trajectòria.

L'ús de la predicció per situar la finestra de seguiment de forma més òptima permet a la vegada definir la possibilitat de una extrapolació del moviment en qualsevol dels següents casos:

- si la característica local seguida es perd de forma temporal (possible oclusió), o no ha sigut detectada.
- si tot i haver-se realitzar l'aparellament, aquest no supera alguna de les restriccions de coherència de trajectòria descrites a l'apartat 3.6.2 (el seguiment no és fiable).

En qualsevol dels anteriors casos, es pren com a posició actual de l'objectiu dins de la imatge el valor de F_k ($P_k = F_k$). La posició tridimensional (X_k) s'estima a partir d'aquest valor i es continua el procés de seguiment durant m mostres. Si al cap de m mostres la característica local no és novament localitzada, el sistema dona per perdut el seu seguiment, i es tanca la seva trajectòria (amb la seva funció coherència associada).

3.6.4. Reinicialització del seguiment

Si durant el seguiment tridimensional de les característiques locals de la imatge, algunes comencen a perdre's, bé perquè surtin de la imatge, resultin ocultes o bé perquè el seu seguiment no ofereixi fiabilitat, es produeix el tancament de la seva trajectòria. En aquest cas, al cap d'un temps de seguiment el conjunt de característiques seguides s'haurà reduït.

Si el nombre de característiques locals seguides baixa per sota d'un mínim es fa necessari una reinicialització del sistema de seguiment. Aquesta operació presenta un alt cost computacional. Al igual que en la fase d'inicialització, cal buscar les característiques locals dins de la imatge esquerra, en les proximitats de les trajectòries perdudes per localitzar-ne de noves. Una vegada localitzades, s'hauran de trobar els seus homòlegs en la imatge dreta, per aconseguir un nou conjunt d'objectius i poder tornar a començar el seguiment.

CAPÍTOL 4

4. Millora del temps de processat

4.1. Introducció

El seguiment d'objectes mitjançant visió per computador necessita d'uns requeriments de maquinari normalment costosos. Amb la idea de crear un sistema ràpid i eficient, però de cost assequible, s'ha implementat un processador d'imatge que opera sobre PC i que proporciona la transformada polar d'aquelles regions de la imatge sol·licitades, la qual cosa permet realitzar el reconeixement i localització de l'objectiu seguit en temps real. Per a la implementació d'aquest processador d'imatge s'ha utilitzat un dispositiu programable (FPGA) que permet la modificació de la topologia de la transformació de la imatge per a adaptar-la a diferents aplicacions.

Els mètodes de seguiment d'objectes mitjançant visió per computador basats en el reconeixement, intenten superar els problemes d'aquells que es basen únicament en una segmentació. El reconeixement aporta més fiabilitat al procés de seguiment mitjançant l'associació de característiques relatives a l'objecte seguit al llarg de la seqüència d'imatges.

El principal inconvenient dels sistemes de seguiment basats en el reconeixement, és que per a cada nova imatge adquirida s'ha de realitzar el reconeixement de l'objecte seguit. Aquesta operació es considerada d'alt nivell dins de la visió per computador, degut a la complexitat intrínseca del problema de la associació de dades i a l'alt cost computacional associat a la seva solució.

La eficàcia d'aquests sistemes de seguiment està limitada per la eficiència del mètode de reconeixement i la seva velocitat d'execució, que determina la latència del sistema. Com que aquests dos paràmetres acostumen a anar contraposats, tots els sistemes de seguiment adopten una solució de compromís entre la viabilitat de l'algoritme de seguiment, la fiabilitat de la detecció i la seva velocitat, limitant normalment el tipus d'objectes que poden ser reconeguts i les circumstàncies en les que poden ser seguits.

Algunes de les solucions presentades per diferents autors intenten millorar la velocitat de processat utilitzant grans equips d'un elevat cost econòmic, la qual cosa limita la seva aplicació de manera generalitzada [Wang, 92] [Fukui, 92] [Inhoue, 93] [Welch, 93] [Ishii, 96].

El sistema de seguiment proposat, descrit al capítol anterior, redueix la quantitat d'informació a processar durant l'associació de dades seleccionant de forma prèvia aquelles regions de la imatge que presenten alguna singularitat (*local features)*. Per a la detecció i identificació d'aquestes singularitats s'ha optat per un mètode basat en la representació polar del contorn de l'objecte.

Aquest mètode de codificació polar permet reduir la representació del contorn de dues dimensions a una. A més a més permet una normalització del la mida, posició i orientació del contorn de manera senzilla (els girs de l'objecte passen a ser una translació a l'espai transformat), per això ha estat utilitzat també per altres autors com a pas previ al reconeixement de formes [Jeng,91] [Sekita,92] [Friedland,92].

En el nostre cas, la codificació polar no es aplicada a la totalitat de l'objecte, sinó d'una forma local a les característiques que presenta el seu contorn [Amat,89][Amat, 92]. La transformació a estat optimitzada i reduïda per a la seva implementació *hardware*. L'àrea transformada es de 15x15 pixels, dels quals s'obtenen 8 radis de 8 valors possibles (Figura 4.1).

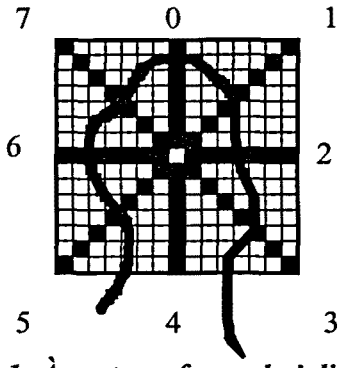


Figura 4. 1. Àrea transformada i distribució dels radis.

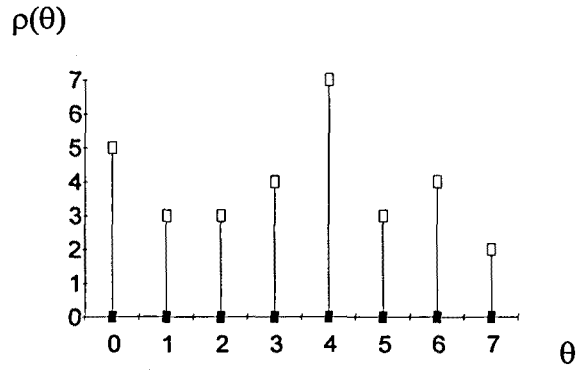


Figura 4. 2. Funció polar discreta del contorn ρ(θ).

Cada radi indica la distància en pixels en què es troba el píxel de contorn de l'objecte en aquesta direcció (Figura 4.2). En cas que hi hagués més d'un píxel de contorn es pren el més proper al centre de l'àrea i en cas que no hi hagi cap píxel de contorn el radi pren el valor màxim (com al radi 4 de la figura 4.1).

Amb l'objectiu de minimitzar el temps de càlcul del sistema de seguiment s'ha optat pel disseny i implementació d'un processador específic que permeti al computador adquirir directament la codificació polar (en pixels) dels contorns existents dins de la regió (o regions) de la imatge que s'està seguint.

4.2. Arquitectura del processador específic implementat

4.2.1. Descripció general del sistema

Per a la implementació del sistema s'ha optat per una arquitectura basada en una tarja processadora d'imatge connectada a un ordinador personal (80386 o superior). (Fig. 4.3). S'ha escollit un PC degut a la relativa alta velocitat i baix preu.

La tarja realitza en temps real i per a cada píxel de la imatge (a *vídeo rate*), la transformació polar descrita anteriorment. El resultat d'aquesta transformació és transferit al PC mitjançant una memòria FIFO. Com que la quantitat d'informació obtinguda per a cada pixel és elevada (24 bits) només es guarden les dades referides a aquelles regions de la imatge (finestres de seguiment) corresponents a la màxima velocitat de desplaçament esperada.

La tarja a més a més, permet que aquestes finestres puguin ser visualitzades sobre la imatge d'entrada a través d'un monitor.

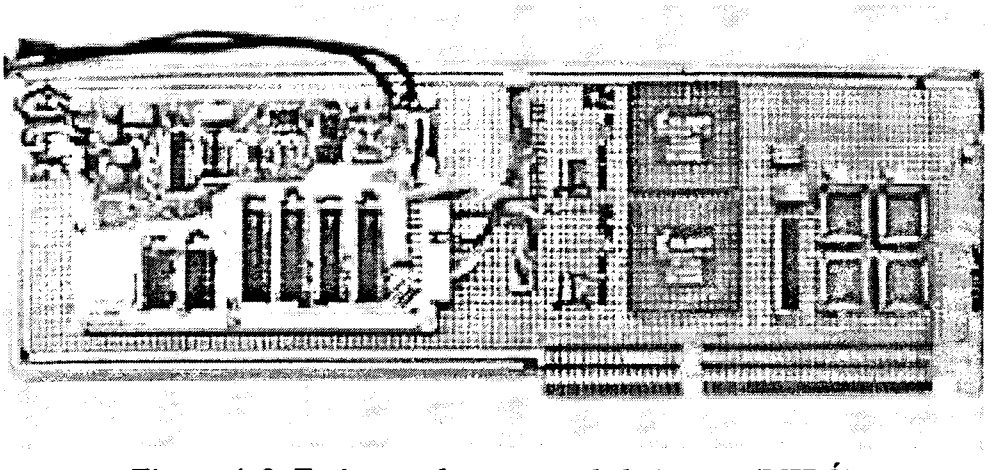


Figura 4. 3. Tarja per al processat de la imatge (MIRÓ).

El processat realitzat pel processador del PC, inclou el procés de reconeixement i localització de l'objectiu per a cada una de les finestres de seguiment, basant-se en les dades obtingudes. Utilitzant la codificació polar obtinguda per a una determinada característica local, es reprograma cada finestra de seguiment, deixant-la preparada per al posterior quadre d'imatge i passa a processar les dades relatives a una altra característica local dins del mateix quadre d'imatge. D'aquesta forma es poden solapar el temps de l'adquisició d'imatge i el temps de l'algorisme de reconeixement (*pipe-line*). S'obté així la cadència total del sistema de 20ms.

El processat d'imatge realitzat per la tarja inclou els següents mòduls, mostrats en l'esquema de la figura 4.4:

- Adquisició i extracció de contorns.
- Memòria de retard de quinze línies d'imatge.
- Circuit de codificació polar.
- Control de la tarja i de les finestres de seguiment.
- Memòria FIFO de comunicació de dades.

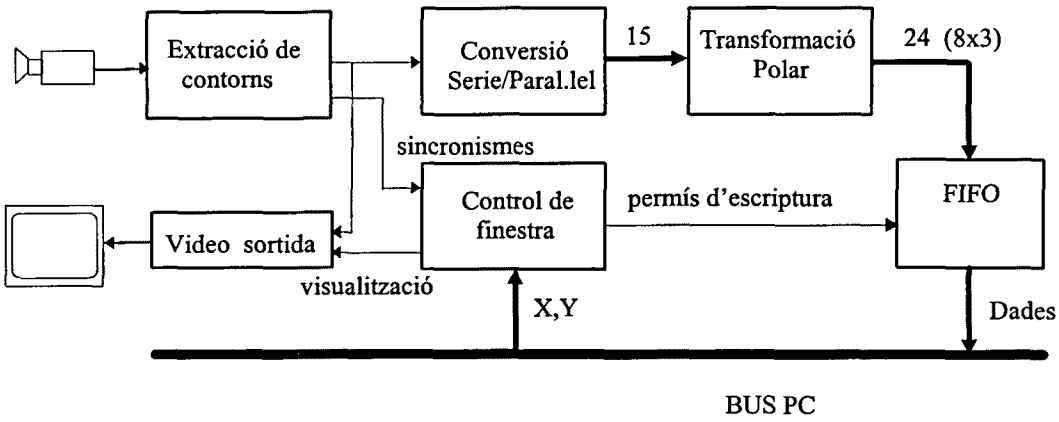


Figura 4. 4. Esquema del processador d'imatge implementat.

Una més completa descripció del *hardware* i de tots els esquemes electrònics del processador implementat es poden trobar en el document de recerca del Departament d'ESAII de la UPC: *"Processador digital d'imatges MIRÓ"* [Aranda,95]. Una segona implementació d'aquest mateix processador es descriu a [Garrido,97].

4.2.2. Adquisició i extracció de contorns de la imatge.

En aquest mòdul es realitza l'amplificació del senyal de vídeo d'entrada, la digitalització de la imatge i l'extracció dels pixels de contorn.

El binaritzat de la imatge sempre presenta el problema de la determinació del valor més idoni del llindar sobre el qual s'ha d'efectuar la separació dels objectes respecte al fons. Mai existeix la garantia que un únic valor pugui realitzar aquesta segmentació dels objectes continguts a la imatge, i per tant, s'ha d'ajustar sempre a les característiques pròpies de l'escena i de la seva il·luminació.

Per aquest motiu s'ha incorporat una etapa de preprocessat consistent en una extracció de contorns utilitzant un operador de 2x2 pixels. La binarització dels gradients obtinguts es realitza comparant amb un llindar variable que permet adaptar-se millor a les característiques pròpies de l'escena. Com és ben conegut, una imatge de contorns binària presenta una estabilitat molt elevada als canvis d'il·luminació i ofereix molta més informació que una imatge binaritzada al mateix cost (1 bit per pixel).

A més a més, dins d'aquest mòdul es realitza la extracció del sincronisme de quadre i de línia del senyal i es genera el rellotge de freqüència de mostreig que a més a més actuarà com a rellotge general del sistema. Aquests senyals es faran servir com senyals de control de posició del píxel en curs i per al control de posició de la finestra de seguiment. El rellotge del sistema ha estat ajustat a la freqüència de 7,37 MHz per aconseguir píxel quadrat, aspecte important per a poder garantir la invariabilitat a la rotació de la transformació polar del contorn dels objectes seguits. Aquesta freqüencia proporciona una resolució de la imatge de 383x287 pixels.

Aquest mòdul incorpora, així mateix, una sortida de vídeo per un monitor en el qual es superposen el senyal d'entrada (o la imatge de contorns) amb els sincronismes i amb una visualització de la finestra de seguiment, la forma de la qual pot ser definida per l'usuari a través del programa.

L'etapa d'extracció de contorns es pot substituir per una binarització en aquelles aplicacions industrials on les condicions d'il·luminació siguin estables i els objectes estiguin ben contrastats respecte el fons. Això reduiria el cost econòmic del processador implementat.

De fet, aquesta etapa de preprocessat pot ser substituïda per d'altres plaques *hardware* desenvolupades al nostre departament que realitzen la segmentació dels objectes utilitzant diferents tipus d'informació. Entre d'altres exemples poden citar la segmentació per color [Batlle,93], la segmentació per diferència [Aranda,94] o la segmentació per textures [Grau,97]. També es possible realitzar un preprocessat amb una segmentació multi-paramètrica disposant aquests preprocessadors en paral·lel o en cascada (*pipe-line*). L'única restricció és que el resultat de tal segmentació sigui una imatge binària prou robusta al soroll i de pixel quadrat.

4.2.3. Memòria de retard

Donat que la imatge es adquirida en sèrie, píxel a píxel, és necessari una nova conversió sèrie/paral·lel de la imatge per poder alimentar el circuit de codificació polar amb columnes d'imatge de 15 pixels. Aquest subministrament ha de sincronitzar-se per a cada nou píxel procedent del mòdul d'adquisició i extracció dels contorns.

Habitualment aquesta conversió és implementada mitjançant registres de desplaçament de longitud igual a una línia que s'alimenten en sèrie. Degut a l'alt nombre de registres que serien necessaris per fer el retard de 15 línies d'imatge, en aquesta tarja processadora, la conversió ha estat realitzada amb una memòria RAM estàtica de 16 bits amb E/S separada, on cada columna de la imatge ocupa una adreça

física. A cada nou cicle de rellotge la adreça es incrementada i una nova columna de 15 pixels es tractada. Durant el primer semicicle la columna de pixels és llegida i enviada al circuit de codificació polar. Durant el segon semicicle s'escriu sobre la mateixa adreça de memòria, la mateixa columna desplaçada i alhora incorporant el nou píxel mostrejat. D'aquesta forma les dades queden preparades per la següent lectura.

4.2.4. Codificació polar de la imatge

La codificació polar de la imatge es realitza utilitzant una matriu de 15x15 biestables que contenen la regió que ha de ser transformada. La informació arriba en una columna de 15 pixels que és desplaçada dins de la matriu a cada nou cicle de rellotge. A cada nou pixel, i per tant a *video rate*, es disposa d'una regió 15x15 de la imatge centrada en el pixel en curs.

La seva implementació ha estat realitzada sobre una XC3064-70 FPGA de XILINX, que ofereix com a avantatges respecte d'altres possibilitats (com la implementació en VLSI, també estudiada) la seva versatilitat i rapidesa de disseny. Per a la seva programació ha estat utilitzat el programari de desenvolupament DASH de FutureNet, basat en una descripció gràfica del circuit. No ha estat necessària la utilització de descripcions més abstractes com les requerides per VHDL.

Les sortides dels biestables seleccionats són agrupades en 8 radis de 7 pixels al voltant del píxel central de l'àrea tractada. (fig. 4.1). Cada grup de pixels és dirigit com entrada a un codificador de prioritat, ordenats segons la seva distància al píxel central. D'aquesta manera s'obté una paraula de 3 bits per cada radi, que indica la distància al centre del primer píxel de contorn que talla el radi en qüestió. El valor 111 (7) ha estat reservat per codificar la possibilitat que cap dels pixels del radi talli un pixel de contorn (radi considerat de longitud infinita).

En total aquest processat proporciona a cada nou cicle 24 (8x3) bits de sortida per a cada pixel de la imatge. Aquests 24 bits contenen la informació rellevant, $\rho(\theta)$, per al reconeixement de la regió de 225 (15x15) pixels que envolta al pixel tractat. Aquesta codificació comporta, per tant, una reducció de gairebé 1:10 en quantitat de bits, mantenint prou informació per poder fer la selecció inicial de les característiques locals presents en la imatge i el seu posterior reconeixement durant l'aparellament estereoscòpic i el seguiment.

Un dels avantatges del disseny sobre FPGA resideix en que aquesta distribució de radis i codificació poden ser modificades per al seguiment d'algun objecte específic (o fins i tot d'altres aplicacions a part del seguiment) sense necessitat de modificar les connexions de la tarja [Amat,93].

De tota la informació proporcionada per aquesta codificació només es guarda aquella relativa als pixels que estan inclosos dins de les regions definides per l'usuari (finestres de seguiment). Si el píxel en curs, que ocupa la posició central de la regió tractada, es dins d'una finestra de seguiment, llavors la codificació obtinguda es desa a la FIFO de sortida on pot ser llegida pel processador del PC.

4.2.5. Control de la tarja i de la finestra de seguiment

Aquest mòdul conté la descodificació de les adreces de control, el registre de control d'estat, el control de posició de la finestra de seguiment i la lògica de control de L/E de les memòries FIFO utilitzades.

Aquest mòdul també va ser implementat utilitzant la mateixa FPGA, i utilitzant el mateix programari de desenvolupament DASH. Utilitzant el model XC3064-70, es va arribar fins al 98% d'ocupació. Per evitar aquesta saturació, en una versió posterior (fig. 4.3) aquest mòdul ha estat implementat en una segona FPGA (més una PAL per la descodificació d'adreces). D'aquesta forma es deixa prou espai lliure a les dues FPGAs com per poder implementar algunes ampliacions de les capacitats del processador presentat (veure el capítol de línies futures).

4.2.5.1. Adreces de control i dades

La tarja processadora d'imatges es comunica amb el PC mitjançant el seu propi bus ISA de 16 bits. Per a la comunicació s'han fet servir sis adreces de memòria que poden ser modificades mitjançant un conjunt de microinterruptors:

- Dues adreces per llegir el contingut de las memòries FIFO.
- Dues més per programar la posició (X,Y) de la finestra de seguiment.
- Una per llegir el registre d'estat de la placa (consistent bàsicament en informació utilitzada per al test del processador durant la seva implementació).
- Una per escriure el registre de control d'estat.

De fet, com es pot observar, amb tres adreces físiques n'hi ha prou si es diferencia la lectura/escriptura amb el corresponent bit de control.

Per transferir les dades contingudes a les memòries FIFO (una paraula de 24 bits per pixel) a la memòria del PC, són necessaris dos accessos a diferents adreces, donada la limitació en l'amplada del bus de dades del bus ISA (16 bits). Això ralenteix el procés de lectura tot i que no és tracta d'un temps crític (Veure al final del capítol els resultats aconseguits).

En el cas de la programació de la finestra de seguiment passa el mateix (tot i no ser tan dramàtic per tractar-se d'un sol accés per finestra). La nova posició (X,Y) és una paraula de 18 bits (9 + 9), i per tant, amb un bus ISA és obligat realitzar dos accessos a memòria per transferir aquesta informació.

D'aquests comentaris es resolt que una de les millores seria la implementació de la tarja processadora, per poder operar sobre bus PCI.

4.2.5.2. Registre de control

Mitjançant el registre de control es poden programar les següents funcions:

- El tipus de visualització al monitor de les finestres de seguiment (transparent, creu, marc, àrea).
- La mida de la finestra de seguiment (8x8, 16x16, 32x32, 256x256).
- El permís d'escriptura de la FIFO.
- L'esborrat de la FIFO.

Els diferents tipus de visualització de les finestres de seguiment sobre el monitor connectat a la tarja poden ser diferents per a cada finestra. També es permet una combinació dels tipus per aconseguir de nous (creu més marc, per exemple). Això proporciona una identificació clara dels objectius per part de l'usuari (o programador del sistema).

Cada finestra de seguiment, de forma independent, permet ser programada d'una mida concreta (8x8, 16x16 ó 32x32). Quan més petita és la finestra més ràpidament es poden processar les seves dades. El número de dades creix ràpidament si dupliquem la dimensió (64, 256 i 1K paraules respectivament). Es per això que es permet el canvi de dimensió de forma dinàmica en funció de la velocitat (i acceleració) que presenta la característica local seguida (veure l'apartat 3.6.4.).

La mida 256x256 està pensada per gravar la codificació polar cada 8 pixels, tant en la direcció horitzontal com en la vertical, degut per una banda a la limitació en la mida de la FIFO i per l'altra en el temps necessari per poder transferir totes aquestes dades (64Kparaules de 24 bits impliquen 128Kaccessos). Es graven per tant, únicament, 1Kparaules (corresponen a la mateixa quantitat d'informació que una finestra de seguiment de 32x32 pixels).

S'obté d'aquesta forma un entramat de codificacions ràpid de llegir, que és molt útil per a la selecció i localització inicial de les característiques locals de l'objecte que poden ser seguides. Com que la transformació polar es realitza en una regió de 15x15 pixels, queda garantit el solapament de les codificacions assignades a cada pixel de l'entramat, de forma que es pot afirmar que aquest entramat escombra completament tota l'àrea de la finestra seleccionada (tot i que no té un error de resolució igual a 8 pixels).

Aquesta finestra de 256x256 permet així una ràpida inicialització de l'algoritme de seguiment sense la necessitat d'informació inicial sobre la posició de l'objecte a seguir. També en cas de pèrdua de l'objectiu (o objectius) durant el seguiment, pot ser utilitzada en un últim intent de trobar la seva localització dins de la imatge o per obtenir d'altres objectius nous a seguir (reinicialització). Aquestes possibilitats queden millor definides als apartats 4.2.7. i 4.2.8.

4.2.5.3. Posició de la finestra de seguiment

Per al control de la posició de la finestra de seguiment s'utilitzen dos comptadors que proporcionen la coordenada del píxel en curs. Aquesta posició és comparada amb la programada per l'usuari als registres X,Y, la qual indica l'inici de la finestra. El final de la finestra ve determinat per la mida de la finestra definida al registre de control. La finestra de seguiment determina l'àrea de la imatge que ha de ser guardada a la FIFO.

Al finalitzar la finestra de seguiment i per tant el magatzematge de dades a la FIFO, la tarja proporciona al microprocessador una interrupció que és utilitzada com avís de final d'adquisició de la transformacions polars al voltant de la característica requerida.

D'aquesta manera el microprocessador pot reprogramar la finestra de seguiment a una altra zona de la imatge, fins i tot al mateix quadre sempre que sigui en línies inferiors. Així, si s'ordenen les finestres, es disposa de més d'una finestra per quadre d'imatge sense necessitat de replicar el *hardware.*

En el cas que dues finestres es solapin parcialment o comparteixin línies, la programació de les seves adquisicions es realitza de forma alternada, fent una mostra cada dos quadres. Per a la majoria d'aplicacions aquesta freqüència de mostreig és suficient (25 mostres/segon). Això fa que no es consideri rentable la replicació del control de posició de finestra a nivell *hardware.*

4.2.6. Memòria de dades

Aquest mòdul ha estat implementat amb una memòria FIFO de 2Kbytes (ampliables a 4Kb i 16Kb amb el mateix encapsulat) on es guarda la imatge una vegada transformada i d'on és llegida pel PC. Com ja s'ha especificat, no es guarda la codificació polar obtinguda sobre tota la imatge, sinó únicament dins l'àrea continguda a les finestres de seguiment, amb el conseqüent estalvi de memòria i de temps d'adquisició de dades.

Aquesta memòria realitza la funció de pulmó, donada la diferència de velocitat entre la escriptura i la lectura de les dades, i resol a baix cost el problema del doble accés per part del codificador que escriu sobre la memòria i el PC que la llegeix de forma asíncrona, sense necessitat de circuits addicionals de sincronització [Martinez,92].

El banc de memòria està constituït per 4 blocs físics (de 9 bits) i dos blocs lògics (de 16 bits). Aquesta divisió ha estat necessària degut a la limitació de 16 bits imposada pel bus a la transferència de dades. Com les dades a transferir són paraules de 24 bits ha estat necessari separar la informació en dos bancs. La divisió no ha sigut pas arbitrària, separant els quatre radis de direcció cartesiana del quatre diagonals. Aquesta divisió permet una optimització dels algoritmes d'associació de dades que no precisen de tota la informació relativa a l'objecte.

4.2.7. Inicialització del sistema

La fase d'inicialització per situar les finestres de seguiment sobre les característiques a seguir, pot ser programada de diferents formes. Es proposen tres:

- *Inicialització manual:*

En que un usuari selecciona manualment els elements de l'escena que cal seguir i inicialitza el seguiment quan considera oportú. Aquest mètode és vàlid per aplicacions de teleoperació.

- *Inicialització preprogramada:*

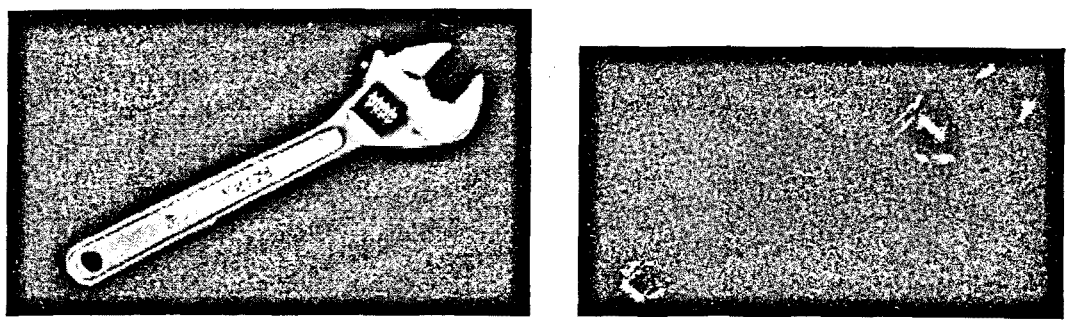
S'indica al sistema la zona on apareix l'objecte a seguir, així com la seva codificació polar (patró). Aquesta forma pot ser aplicada en aquelles aplicacions ben conegudes a priori on també són ben conegudes les característiques locals que es volen seguir.

- *Inicialització automàtica:*

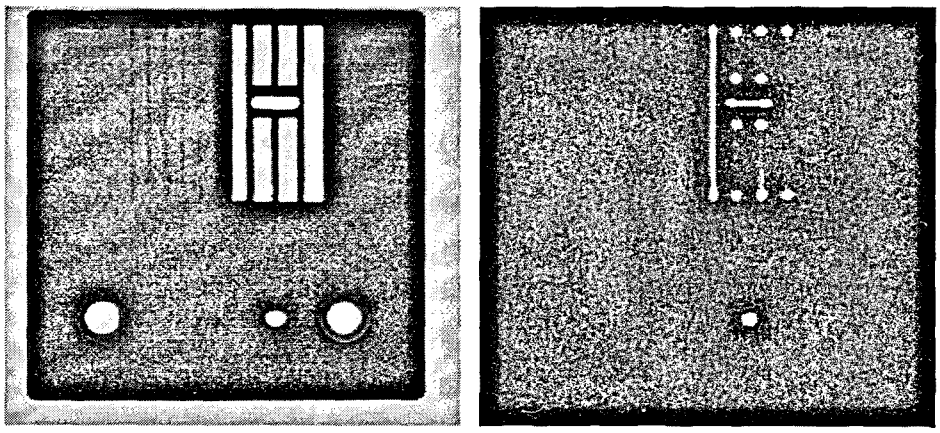
El sistema és programat per fer un escombrat per la imatge (zona de búsqueda de 256x256) i localitzar en un sol quadre (20 ms) aquelles posicions de la imatge que presenten alguna singularitat (seguint el criteri que va ser descrit a l'apartat 3.3, Eq. 3.4). En una segona passada es demana informació al voltant d'aquelles posicions obtingudes anteriorment (utilitzant finestres de 16x16). D'aquesta forma es localitza de forma precisa (resolució igual a un pixel) les posicions i les transformacions polars (descriptors) inicials de les característiques locals dels elements relevants de la imatge.

Amb aquest mètode s'aconsegueix una inicialització ràpida (menys de 100 ms) i automàtica del sistema de seguiment ja que els patrons a seguir i la seva posició poden ser obtinguts directament de la imatge transformada.

Això presenta un clar avantatge respecte als sistemes de seguiment basats en la adquisició de la imatge mitjançant una placa convencional, on el mateix procediment pot trigar alguns segons si es realitza sobre tota la imatge.



a) Característiques locals localitzades sobre una clau anglesa.



b) Característiques locals localitzades sobre una peça industrial.

Figura 4. 5. Localització automàtica de les característiques locals més fiables.

4.2.8. Algorisme de seguiment

Durant la fase de seguiment altres sistemes acostumen a buscar el patró seguit al voltant de la darrera posició en la qual va ser localitzat o d'una estimació de la posició prevista.

Al sistema presentat aquesta estimació de la posició no és necessària, ja que les dades relatives a la finestra de seguiment han estat gravades de manera seqüencial a la FIFO per la pròpia tarja, la qual cosa implica un estalvi d'operacions en l'accés a la informació. Aquesta informació consistent en la codificació polar de la imatge respecte a cada píxel de la finestra de seguiment, és comparada amb la codificació polar buscada (patró).

L'algoritme de reconeixement està basat en la funció de comparació (Eq. 3.8):

$$D(i,j) = \sum_{\theta=0}^{7} \left[r(\theta) - m_{ij}(\theta) \right]^2 \quad \forall (i,j) \in \textit{finestra de seguiment}$$

Aquesta funció d'error és obtinguda entre la codificació polar del patró, $r(\theta)$, i aquella corresponent a cada píxel (i,j) de la finestra de seguiment, $m_{ij}(\theta)$. Així, el valor mínim d'aquesta funció determina la nova posició de la finestra de seguiment.

La funció no és calculada de manera algebraica sinó que tots els possibles resultats estan tabulats a la RAM del PC per blocs, la qual cosa accelera el procés de comparació. Cada codificació polar, $r(\theta)$ *i* $m_{ij}(\theta)$, consta de 24 bits d'informació (8 radis amb 3 bits cadascun). Per tabular tots els possibles resultats de la funció distància *D*, caldria una memòria de 2^{48} paraules de 9 bits (cal tenir en compte que el valor màxim de la funció distància pot ser $8*7^2 = 396$).

El que s'ha fet és comparar els radis de les dues codificacions de dos en dos. Així la memòria necessària a quedat reduïda a 2^{12} adreces = 4Kbytes (ara el valor màxim de la funció distància només pot arribar a $2*7^2 = 98$). Ara, calen 4 accessos a aquesta taula per conèixer el valor de comparació dels 8 radis de la codificació. Finalment, cal sumar els resultats parcials obtinguts.

Amb aquesta estratègia de càlcul per taula afavoreix poder operar amb valors no lineals al mateix cost. Per exemple ha sigut provada la penalització de la comparació entre radis quan el seu valor és 111 (codificació per infinit, o no trobat).

El càlcul algebraic de la funció distància requereix de 23 operacions aritmètiques: 8 restes, 8 productes i 7 sumes. La tabulació de resultats parcials proposada redueix a 7 les operacions: 4 accessos a memòria i 3 sumes. Aquesta reducció (1/3) és important si es té en compte l'elevat nombre de comparacions que s'ha d'efectuar.

Evidentment quant més petit és el desplaçament de l'objecte entre quadres d'imatge mesurat en píxels, més petita pot ser la finestra de seguiment i més ràpid el procés de comparació. L'usuari pot programar finestres de 8x8,16x16, 32x32 de forma dinàmica en funció de la velocitat (i acceleració) esperada de l'objecte. El sistema també permet realitzar aquesta estimació de dimensió de les finestres de seguiment de forma automàtica (veure l'apartat 3.6.4).

El principal problema dels algoritmes de seguiment d'objectes resideix en la possible rotació d'aquests entre imatges consecutives. Una petita rotació del patró seguit provoca un error molt elevat ja que és vist com un canvi de forma per l'algoritme de reconeixement.

Per resoldre aquest problema es realitza una triple comparació de la codificació obtinguda $m_{ij}(\theta)$ amb les funcions $r(\theta)$, $r(\theta-1)$ i $r(\theta+1)$. Les dues darreres corresponen al mateix patró girat 45° en ambdós sentits. És aquí on la representació polar del contorn de l'objecte ofereix una millora important respecte a d'altres tècniques, ja que la rotació del patró passa a ser un canvi d'índex en l'accés a la seva informació. Si l'error mínim correspon amb algun dels patrons girats aquest passa a ser el nou patró. D'aquesta manera s'aconsegueix que la velocitat de rotació de l'objecte pugui ser de fins a 45° per quadre d'imatge.

4.3. Resultats

Per poder efectuar una avaluació de la millora en el temps de processat que aporta el processador d'imatge descrit a l'apartat anterior, es presenta una comparació amb el temps de processat necessari en el cas d'utilitzar una tarja comercial d'adquisició i processat d'imatge.

La comparació s'ha realitzat concretament amb la tarja MVP-AT de Matrox. L'elecció d'aquesta tarja es deu a que és de característiques adients per efectuar el seguiment estereoscòpic (disposa de quatre canals d'entrada no multiplexats) i té una bona relació preu/prestacions. És a més a més, una placa ben coneguda internacionalment. Tots els mòduls descrits al capítol tres, han estat inicialment programats i optimitzats per treballar sobre aquesta tarja comercial. Es disposa per tant de dades concretes relatives a temps de processat del que podem anomenar la solució *software*.

Els principals inconvenients d'aquesta tarja comercial per aplicacions de seguiment, radica en l'elevat temps necessari per fer l'adquisició de la imatge (que oscil·la entre 60ms i 80 ms) i la transferència a la memòria RAM del PC de les dades relatives a una finestra d'interès. Tot i que recentment han aparegut en el mercat targes més ràpides (basades en el nou BUS PCI), i que permeten fer adquisició continua (mitjançant una memòria *ping-pong* ubicada en molts casos sobre la pròpia RAM del PC), l'adquisició i la transferència de dades continuen sent els principals problemes.

Per poder fer la comparació dels resultats obtinguts operant amb diferents targes, s'han separat els temps d'adquisició, el temps de transferència de les dades i el temps de processat de la informació. En el cas de la tarja MVP el temps de processat inclou la codificació polar de la imatge. Aquesta part del processat és directament proporcionada per la tarja específica implementada (Miró) i per tant no està inclosa en el seu temps de processat de dades. Tots els temps de processat que apareixen a les taules corresponen a un PC 486 a 100MHz.

4.3.1. Seguiment d'una característica local

Cal aclarir que l'adquisició de la imatge amb la tarja MVP es pot realitzar en paral·lel amb el processat de les dades, però no amb la transferència d'aquestes. Només es poden solapar, per tant, els temps d'adquisició i processat. La tarja "Miró" sí permet l'accés a les dades durant l'adquisició (gràcies al *buffer* de memòries FIFO), permetent el solapament dels temps de transferència i processat amb el temps d'adquisició.

El temps de processat depèn de la dimensió de la finestra de seguiment.

finestra de 32x32	Adquisició	Transferència	Processat	Temps total
MVP	60ms	20ms	140ms	160ms
Miró	20ms	3ms	6ms	20ms

finestra de 16x16	Adquisició	Transferència	Processat	Temps total
MVP	60ms	5ms	35ms	65ms
Miró	20ms	0,75ms	1,5ms	20ms

finestra de 8x8	Adquisició	Transferència	Processat	Temps total
MVP	60ms	1,25ms	8,75ms	61,25ms
Miró	20ms	0,18ms	0,38ms	20ms

Un PC equipat amb el processador específic implementat necessita un temps de processat molt inferior a un quadre d'imatge (20ms) per efectuar el seguiment d'una característica local. Per tant, podem obtenir resultats a cada quadre, es a dir a 50Hz.

4.3.2. Seguiment de *n* característiques locals

És en el seguiment de múltiples regions singulars de la imatge, on la tarja "Miró" mostra una aportació més rellevant. Si es pren com a exemple finestres de 16x16, tenim que per *n* característiques locals els temps de processat són:

finestra de 16x16	Adquisició	Transferència	Processat	Temps total
MVP	60ms	5·n ms	35·n ms	40·n ms ($\forall n \geq 2$)
Miró	20ms	0,75·n ms	1,5·n ms	20 ms (si n <9) (*) 2,25·n ms (si n ≥9)

(*) Cal recordar que en el cas de la placa "Miró", el nombre de quadres d'imatge necessaris per adquirir les dades coincideix amb el nombre de finestres màxim que hi hagi solapades per cada línia d'imatge. Així si per exemple una línia en té dues finestres solapades i una altra en té tres, el temps d'adquisició serà de tres quadres (60ms). En el pitjor cas caldrien tants quadres com finestres de seguiment es pretenen adquirir. Aquest cas però, és molt improbable ja que implica que totes les finestres de seguiment comparteixen al menys una línia d'imatge.

per diferents valors de *n* tenim:

n finestres	n=2	n=4	n=8	n=16	n=100
MVP	80ms	160ms	320ms	640ms	4000ms
Miró (*) (millor cas)	20ms	20ms	20ms	36ms	225ms
Relació	x4	x8	x16	x17'7	x17,7
Miró (**)	20ms (1)	40ms (2)	60ms (3)	80ms (4)	400ms (20)
nF/Q	2	2	2,6	4	5
Relació	x4	x4	x5,3	x8	x10

(*) En el *pitjor cas* (solapament de totes les finestres de seguiment) el temps de processat amb la tarja Miró és sempre de 20ms·n. *La meitat* que el necessari amb la MVP independentment del nombre de finestres.

(**) Cas més *probable*, en el que s'espera un cert solapament entre finestres (indicat entre parèntesis). S'indica també a la taula el nombre de finestres processades per quadre, nF/Q, per cada cas.

Ha sigut implementat un gestor *software* que s'encarrega d'optimitzar la lectura de les finestres de seguiment que tenen solapament, de forma que fent una reordenació puguin ser llegides en el mínim nombre de quadres possible [Garrido,97]. D'aquesta forma, la gestió de l'adquisició de dades resulta transparent a l'usuari del sistema.

En el cas de finestres de seguiment de 8x8 la probabilitat de solapament disminueix a la meitat i per tant el temps de processat resulta mes petit.

4.3.3. Seguiment estereoscòpic

En el cas de seguiment de dues finestres de forma simultània (corresponents a les imatges esquerra i dreta) el temps d'adquisició es manté ja que es fa en paral·lel, però es duplica el temps de transferència i processat de dades.

El processat de dades necessari per extreure la trajectòria 3D a partir de les dades 2D té el mateix cost per les solucions avaluades, per la qual cosa no resulta necessari afegir el seu temps en la comparació.

Pel cas del seguiment 3D d'una característica tenim (dues finestres de 16x16):

finestra de 16x16	Adquisició	Transferència	Processat	Temps total
MVP	60ms	2·5ms	2·35ms	80ms
Miró	20ms	2·0,75ms	2·1,5ms	20ms

Pel cas general de seguiment 3D de *n* característiques :

finestra de 16x16	Adquisició	Transferència	Processat	Temps total
MVP	60ms	2·5·n ms	2·35·n ms	80·n ms ($\forall$n)
Miró	20ms	2·0,75·n ms	2·1,5·n ms	20 ms (si n <5)(*) 4,5·n ms (si n ≥5)

(*) Sempre que les finestres no es solapin es podran seguir fins a 4 característiques locals en 3D a 50Hz.

per diferents valors de *n* tenim:

n característiques	n=2	n=4	n=8	n=16	n=100
MVP	160ms	320ms	640ms	1280ms	8000ms
Miró (*) (millor cas)	20ms	20ms	36ms	72ms	450ms
Relació	x8	x16	x17,7	x17'7	x17,7
Miró (**)	20ms (1)	40ms (2)	60ms (3)	80ms (4)	460ms (23)
nC/Q	2	2	2,6	4	4,35
Relació	x8	x8	x10,6	x16	x17,4

(*) En el *pitjor cas* (solapament de totes les finestres de seguiment) el temps de processat amb la tarja Miró és sempre de 2·20ms·n . *Quatre vegades inferior* que el necessari amb la MVP independentment del nombre de característiques.

(**) Si suposem com a l'apartat anterior un cert solapament entre finestres estimat com a probable (s'indica entre parèntesis el nombre màxim de finestres per línia a cada imatge). S'indica també a la taula el nombre de característiques processades per quadre, nC/Q, per cada cas.

En el cas de finestres de seguiment de 8x8 la probabilitat de solapament disminueix a la meitat i per tant el temps de processat resulta mes petit.

4.4. Conclusions

La incorporació de la tarja descrita en un PC estàndard, proporciona la capacitat de realitzar el seguiment estereoscòpic en temps real (20ms) de fins a quatre característiques locals d'un objecte present a la escena. Per un nombre prou elevat de característiques locals proporciona al PC velocitats de processat de 8 a 17,7 vegades superiors a les aconseguides amb una tarja d'adquisició d'imatges comercial, degut a la realització per *hardware* de la transformació polar de la imatge.

L'ús d'un dispositiu programable (FPGA) per a l'implementació del processat d'imatge requerit pel sistema, permet fer modificacions en el tipus de transformació de la imatge per a que s'adapti de forma més encertada a la aplicació concreta. En un pla més físic, l'ús d'aquest dispositiu a permès un disseny de la tarja compacte i per tant més fiable, amb un temps d'implementació més curt.

L'elevada velocitat d'operació i el baix cost econòmic del sistema permet la seva utilització en un ampli ventall d'aplicacions industrials, com ara la servoposició de robots. La teleoperació és també una àrea d'aplicació per al sistema, donant al mecanisme teleoperat certa autonomia.